Rik Leemans *Editor*

Ecological Systems

Selected Entries from the Encyclopedia of Sustainability Science and Technology

Springer

Ecological Systems

This volume collects selected topical entries from the *Encyclopedia of Sustainability Science and Technology* (ESST). ESST addresses the grand challenges for science and engineering today. It provides unprecedented, peer-reviewed coverage of sustainability science and technology with contributions from nearly 1,000 of the world's leading scientists and engineers, who write on more than 600 separate topics in 38 sections. ESST establishes a foundation for the research, engineering, and economics supporting the many sustainability and policy evaluations being performed in institutions worldwide.

Editor-in-Chief
ROBERT A. MEYERS, RAMTECH LIMITED, Larkspur, CA, USA

Editorial Board
RITA R. COLWELL, Distinguished University Professor, Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD, USA
ANDREAS FISCHLIN, Terrestrial Systems Ecology, ETH-Zentrum, Zürich, Switzerland
DONALD A. GLASER, Glaser Lab, University of California, Berkeley, Department of Molecular & Cell Biology, Berkeley, CA, USA
TIMOTHY L. KILLEEN, National Science Foundation, Arlington, VA, USA
HAROLD W. KROTO, Francis Eppes Professor of Chemistry, Department of Chemistry and Biochemistry, The Florida State University, Tallahassee, FL, USA
AMORY B. LOVINS, Chairman & Chief Scientist, Rocky Mountain Institute, Snowmass, USA
LORD ROBERT MAY, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
DANIEL L. MCFADDEN, Director of Econometrics Laboratory, University of California, Berkeley, CA, USA
THOMAS C. SCHELLING, 3105 Tydings Hall, Department of Economics, University of Maryland, College Park, MD, USA
CHARLES H. TOWNES, 557 Birge, University of California, Berkeley, CA, USA

EMILIO AMBASZ, Emilio Ambasz & Associates, Inc., New York, NY, USA
CLARE BRADSHAW, Department of Systems Ecology, Stockholm University, Stockholm, Sweden
TERRY COFFELT, Research Geneticist, Arid Land Agricultural Research Center, Maricopa, AZ, USA
MEHRDAD EHSANI, Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX, USA
ALI EMADI, Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL, USA
CHARLES A. S. HALL, College of Environmental Science & Forestry, State University of New York, Syracuse, NY, USA
RIK LEEMANS, Environmental Systems Analysis Group, Wageningen University, Wageningen, The Netherlands
KEITH LOVEGROVE, Department of Engineering (Bldg 32), The Australian National University, Canberra, Australia
TIMOTHY D. SEARCHINGER, Woodrow Wilson School, Princeton University, Princeton, NJ, USA

Rik Leemans
Editor

Ecological Systems

Selected Entries from the Encyclopedia of Sustainability Science and Technology

Springer

Editor
Rik Leemans
Environmental Systems Analysis Group
Wageningen University
Wageningen 6700 AA,
The Netherlands

This book consists of selections from the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers, originally published by Springer Science+Business Media New York in 2012.

ISBN 978-1-4614-5754-1 ISBN 978-1-4614-5755-8 (eBook)
DOI 10.1007/978-1-4614-5755-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012954278

© Springer Science+Business Media New York 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Ecological Systems, Introduction

Rik Leemans

It is estimated that there are roughly 8 million animal species, 600,000 fungi, 300,000 plants, and an unestimated number of microbial species on earth. It is further estimated that only about 25% of the animal, fungi, and plant species have been identified as yet. Further, the interactions among these species and their physical environment are known to an even lesser degree. Meanwhile, the earth's biota faces the prospect of climate change which may be either slow to manifest or extremely rapid, as we pass through a potential tipping point. In addition, human population is set to grow by two billion by 2045 from a present seven billion. This would certainly be rated as a population explosion, and these numbers together with expected (and hoped for) economic growth will stress the entire ecology of the earth. The question is how to cope with all of the above. That is actually the subject and goal of this entire *Encyclopedia* and our team of well over 700 scientists and engineers, and also the subject of this section from an ecological systems viewpoint. Our objective in this section is to provide a significant portion of the scientific and engineering basis of the systems ecology of the earth and to provide this in 14 detailed entries written at a level for use by university students through practicing professionals. Our approach is to recognize that we cannot wait for a complete data set for the biota before assessing and planning and acting to preserve the ecological balance of the earth. Therefore, methodology is presented aimed at identifying the key interactions and environmental effects and enabling a systems level understanding even at our present state of factual knowledge.

For, ease of use by students, each entry begins with a glossary of terms, while at an average length of 20 print pages each, sufficient detail is presented for utilization

---

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

R. Leemans (✉)
Environmental Systems Analysis Group, Wageningen University,
Wageningen 6700, The Netherlands
e-mail: rik.leemans@wur.nl

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_1,

by professionals in government, universities, and industry. The reader is also directed to the closely related sections: Earth System Monitoring, Introduction and also Transport and Fate of Chemicals in the Environment, Introduction.

Each of the entries is summarized below.

Ecosystems, Adaptive Management – Adaptive management is an approach to natural resource management that emphasizes learning through management based upon the philosophy that knowledge is incomplete and much of what is thought to be known is actually wrong, but despite uncertainty managers and policymakers must act. This entry includes: Structure Decision Making, Participatory Active Adaptive Management, Adaptive Governance, and Adaptive Management and Law.

Marine Biogeochemistry – The biogeochemistry of the world oceans has been studied for many decades, and major advances in understanding have been linked with development of new techniques and tools that allow the accurate representation of various organic and inorganic pools within the water. It has been shown that some critical bioactive compounds (containing the elements: carbon, nitrogen, phosphorus, oxygen) occur with particular atomic ratios to one another that are relatively invariant over space and time pools. The following cycles are presented in detail: carbon ocean Acidification, nitrogen and phosphorous, iron (including iron ocean fertilization), sulfur, and oxygen.

Atmospheric Biogeochemistry – Biogeochemistry represents the interaction of biology, chemistry, and geology in the Earth system. For many processes, an understanding of biological uptake and emission, chemical processing, and geological sequestration is necessary to resolve the sources and sinks of a particular constituent. For example to discover the sources and sinks of atmospheric carbon dioxide, it is important to understand how biota take up carbon dioxide and chemically convert the carbon to organic carbon, and then how this organic carbon is used either to produce energy by biota or is deposited to the land or ocean surface and can become sequestered in geological formations. Often when biogeochemistry is referred to, one refers to the nutrient cycling of important nutrients. Atmospheric biogeochemistry, as defined in its narrowest sense, is the flux of nutrients and pollutants important for biogeochemistry through the atmosphere. In its broadest sense, atmospheric biogeochemistry is the interaction of the atmosphere and atmospheric processes with the biogeochemistry of the earth system. This broader definition is used here.

Ecological Succession and Community Dynamics – Ecological succession and the dynamic change of communities is one of the pillars of modern ecology and it has considerable influence on sustainability science. The processes that drive succession and the regularly of patterns of succession have been the topic of considerable debate over the past century – due in no small part to the importance of the concept to basic and applied ecology. Succession is at its simplest level the pattern of change in ecological communities and these changes can be read on the landscape if one knows regional history or conversely can be used to interpret a landscape’s history.

Ecosystem Engineers, Keystone Species – This contribution focuses on two ecological phenomena. The first is "keystone species" which is defined as a species (mostly of high trophic status) whose activities exert a disproportionate influence on the patterns of species occurrence, distribution, and density in a community. The second is the concept of "ecosystem engineers" defined as organisms that directly or indirectly modulate the availability of resources (other than themselves) to other species by causing physical state changes in biotic or abiotic materials. Models have shown that the loss of species with a large number of trophic interactions can trigger high numbers of secondary extinctions with serious consequences for species persistence. The importance of keystone species can also be strongly linked with ecosystem engineering. For example, the large impact of sea otters in kelp forest ecosystems results from the coupling between engineering effects and a trophic cascade.

Ecosystem Services – This entry provides an overview of issues concerning the identification, biophysical and economic characterization, and safeguarding of ecosystem services. The concept of ecosystem services has a long written history, reaching back at least as far as Plato. This review history focuses on recent advances such as the Millennium Ecosystem Assessment, and advances in spatial modeling, economic valuation, and policy implementation. Examples of novel finance and policy mechanisms, including water funds, marine spatial planning, land-use planning and human development, and global policy efforts are presented. The entry concludes with a discussion of the largest research and implementation challenges in this field, identifying the issues that will frame the future growth of the concept of ecosystem services.

Ecosystems and Spatial patterns – Ecological processes such as forest disturbances act on ecosystems at multiple spatial and temporal scales to generate complex spatial patterns. These patterns in turn influence ecosystem dynamics and have important consequences for ecosystem sustainability. Analysis of ecosystem spatial structure is a first step toward understanding these dynamics and the uncertain interactions among processes. Quantitative characterization of spatial patterns and their rates of change in natural environments are essential to understanding ecological processes and to inform sustainable management techniques that aim to minimize degradation and alteration of ecosystem dynamics. This entry covers: sources of heterogeneity, spatial analysis, spatial analyses within the ecosystem and spatial analyses among ecosystems.

Ecosystem Flow Analysis – It is necessary to understand the ecosystem as a thermodynamic system and how the energy flows enter, interconnect, and disperse from the environmental system. Ecological network methodologies exist to investigate and analyze these flows. In particular, partitioning the flow into boundary input, non-cycled internal flow, and cycled internal flow shows the extent to which reuse and recycling arise in ecosystems. The intricate, complex network structures are responsible for these processes all within the given thermodynamic constraints. Design of sustainable human systems could be informed by these organizational patterns, in order to use effectively the energy available. This entry demonstrates the need for flow analysis, provides a brief example using a well-studied ecosystem, and discusses some of the ecosystem development tendencies which can be addressed using ecosystem flow analysis.

Ecotones and Ecological Gradients – Ecotones are areas of steep transition between ecological communities, ecosystems, or ecological regions (such as Mediterranean and desert). Ecotones often occur along ecological gradients. Such gradients are created as a result of spatial shifts in elevation, climate, soil, and many other environmental factors. Ecotones commonly coincide with areas of sharp climatic transition along environmental gradients. They occur at multiple spatial scales, from continental-scale transitions between major biomes to small-scale ecotones where local vegetation communities and microhabitats coincide. They show a diversity of boundary types that range from natural boundaries (e.g., altitudinal, latitudinal transitions) to human-generated ecotones (e.g., forest clear-cut edges or urban ecotones). Ecotones have been studied in the past four decades in an ecological context and in recent years are receiving increasing attention in the context of biodiversity conservation. Various studies have shown that species richness and abundances tend to peak in ecotonal areas, though exceptions to these patterns occur. Ecotones are “natural laboratories” for studying a range of evolutionary processes, such as the process by which new species form, also termed speciation. Recent evidence suggests that ecotones may serve as speciation centers. Some researchers argue that ecotones deserve high conservation investment, potentially serving as speciation and biodiversity centers. Because ecotones are often small in size and relatively rich in biodiversity, conservation efforts in these areas may prove to be an efficient and cost-effective conservation strategy.

Invasive Species – This entry describes the causes and consequences of biological invasions, by synthesizing concepts from population biology, community ecology, evolution, biogeography, and conservation biology. First, the patterns and process of invasion are explored; then, some of its potential ecological and socioeconomic impacts are examined. Some major hypotheses and theoretical concepts explaining patterns of colonization and impact are presented. Next, management approaches to assessing, preventing, and mitigating this problem are considered. The entry ends with a brief glimpse at some of the emerging issues that will likely be the foci of future research.

Landscape Ecology – Landscapes are spatially heterogeneous areas characterized by diverse patches that differ in size, shape, contents, and history. When spatial heterogeneity is considered, scale matters and hierarchy emerges. Landscape ecology is the science and art of studying and improving the relationship between spatial pattern and ecological processes on a multitude of scales and organizational levels. In a broad sense, landscape ecology represents both a field of study and a scientific paradigm. As a highly interdisciplinary and transdisciplinary enterprise, landscape ecology integrates biophysical and analytical approaches with humanistic and holistic perspectives across natural and social sciences. Landscape ecology was initially developed in Europe. With theoretical developments in spatial ecology and technological advances in remote sensing and geospatial information processing, landscape ecology became an internationally recognized field of study in the 1980s. The most salient characteristics of landscape ecology are its emphasis on the pattern-process relationship and its focus on broadscale ecological and environmental issues. Key research topics in landscape ecology include ecological flows in landscape mosaics, land use and land cover change, scaling, relating landscape pattern analysis with ecological processes, and landscape conservation and sustainability.

Species Competition and Predation – Competition and predation are key interactions between species, and are major foci of thought and study in community ecology. They are believed to be major forces structuring natural communities, having critical roles in the determination of species diversity and species composition, and are regarded as important drivers of evolutionary processes. The relationships between the niches of different species determine how they interact through competition and predation, which then have key roles in assembly of local communities, and their reassembly following perturbations. Niches define patterns of linkages between species, their resources, and their natural enemies. These patterns include how linkages change over time, and between different spatial locations, and define mechanisms by which similar species are able to coexist by their effects on competition and predation relationships. The human element in the environment has profound effects on these phenomena. Changing the environment shifts interactions between species, and profoundly modifies the structure of food webs. In the modern day, there is much community reassembly, potentially involving major shifts in competition and predation. Humans transport invasive species that act as predators, prey, and competitors with potentially major effects on the community reassembly process.

Species Diversity Within and Among Ecosystems – Species diversity is a function of species richness, the number of species in a given locality and species evenness, the degree to which the relative abundances of species are similar. Commonly used quantification methods include constructing mathematical indices known as diversity indexes (the Shannon, Simpson, and Margalef indexes being the most widely used) or comparing observed patterns of species abundance to theoretical models. The scale of assessment may range from within a single site or habitat (known as α diversity) to the difference between two or more sites (β diversity); which can then be combined to give γ diversity – the diversity of the landscape. Early research tended to be concentrated on largely α diversity, but there has been a marked increase in studies investigating β and γ diversity during the last decade. The species richness aspect of diversity is often considered the iconic measure of biodiversity since it is well defined and aligns with our intuitive sense of the concept. As the species level is the level at which living organisms are most widely known, species have become the major taxonomic rank for describing biodiversity. Along with populations, it is also the level that most scientists, managers, and policy makers use when referring to biological diversity and it provides a useful means for both monitoring and priority setting. As the species level is one of three fundamental levels of biodiversity (the others are the genetic level and the ecosystem/community level), species diversity is thus an important component of conservation.

Urban Ecology – Within the science of ecology, urban ecology is defined as the study of structure, dynamics, and processes in urban ecological systems. Urban ecology is the study of the relationships of human and nonhuman organisms in urban areas, the interactions of these organisms with the native and built physical environment, and the effects of these relationships on the fluxes of energy, materials, and information within individual urban systems and between urban and nonurban systems. Urban ecology applies the methods and concepts of the biological science of ecology to urban areas, but requires and integrates with the concerns, concepts, and approaches of social sciences to produce a hybrid discipline. Urban ecological systems include individual organisms, populations, communities, and landscapes, as well as buildings and infrastructure. Urban ecology further recognizes specific urban ecosystems as a part of the global biogeochemical, economic, and human demographic system.

Chapter 2
Atmospheric Biogeochemistry

Natalie M. Mahowald

Glossary

Aerosol	A solid or liquid suspended in the atmosphere. The definition usually does not include cloud droplets, although many aerosols have water vapor on their surfaces.
Dry deposition	Removal process for gas and aerosol species in which the species are deposited onto the lower surface due to either turbulent fluxes (overturning air) forcing the constituent to hit and stick to the surface, or from gravitational settling of aerosols. Gravitational settling is the dominant mechanism for removal for larger aerosols.
Lifetime	The atmospheric lifetime of a constituent describes how long the constituent will remain in the atmosphere. It is typically calculated by dividing the total amount of the constituent in the atmosphere by the total flux out of or into the atmosphere. The flux can be due to atmospheric chemical reactions, and/or exchanges between other reservoirs in the earth system (e.g., land or ocean). This lifetime is an e-folding lifetime; if one starts with an initial perturbation of the constituent, the amount of the perturbation remaining after a length of time equal to the lifetime is equal to 1/e of the original value of the perturbation.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

N.M. Mahowald (✉)
Department of Earth and Atmospheric Sciences, Atkinson Center for Sustainable Future,
Cornell University, 2140 Snee Hall, Ithaca, NY 14853, USA
e-mail: mahowald@cornell.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_2,

Wet deposition	Process by which an atmospheric constituent is removed by precipitation. This is especially important for water-soluble and aerosol species.

Definition of the Subject and Its Importance

Biogeochemistry represents the interaction of biology, chemistry, and geology in the Earth system. For many processes, an understanding of biological uptake and emission, chemical processing, and geological sequestration is necessary to resolve the sources and sinks of a particular constituent. For example, to discover the sources and sinks of atmospheric carbon dioxide, it is important to understand how biota take up carbon dioxide and chemically convert the carbon to organic carbon, and then how this organic carbon is used either to produce energy by biota or is deposited to the land or ocean surface and can become sequestered in geological formations. Often when biogeochemistry is referred to, one refers to the nutrient cycling of important nutrients. Atmospheric biogeochemistry, as defined in its narrowest sense, is the flux of nutrients and pollutants important for biogeochemistry through the atmosphere. In its broadest sense, atmospheric biogeochemistry is the interaction of the atmosphere and atmospheric processes with the biogeochemistry of the earth system. This broader definition is used here. The hydrologic cycle is one of the most important biogeochemistry cycles, and is a worthy topic of its own. It is discussed in several other sections of this encyclopedia (e.g., Climate Change and Global Water Sustainability, Water Resource and Quality Management for Adaptation to Climate Change, Water Availability and Quality, Sustainable).

Introduction

Atmospheric biogeochemistry is increasingly relevant for today's society, as evidenced by the changing composition of the atmosphere. Atmospheric biogeochemistry views these changes in the composition as part of a collection of changes to the earth system occurring during the anthropocene. One very serious sustainability issue at the core of atmospheric biogeochemistry is the accumulation of carbon dioxide in the atmosphere. This is a result of human emissions of carbon dioxide both directly from fossil fuel burning, and from land use change, especially tropical deforestation [1]. Interestingly, only about ½ of the carbon dioxide emitted into the atmosphere remains there: The rest is taken up by the land and ocean. This important negative feedback on human emissions of carbon dioxide is a result of biogeochemical cycling of carbon in the system. As human emissions of carbon dioxide grow, it is not clear that the land and the ocean will be able to keep up, and thus more of the carbon dioxide is likely to stay in the atmosphere [2]. Some studies suggest that warming itself leads to less carbon dioxide being taken up by the land and ocean [2], although this also depends on the cycling of other nutrients such as nitrogen as well [3]. Thus understanding the biogeochemistry of carbon is required to understand how much carbon can be emitted in the future.

The atmospheric transport and subsequent deposition of trace gases and aerosols can provide an important mechanism for circulation of different constituents. Whether this deposition serves as a nutrient or a pollutant usually depends on the deposition level, as well as the ability of the biota to process these compounds. Some compounds can be nutrients at low levels, but pollutants at high levels. Thus, referring to the atmospheric transport of nutrients and pollutants as "atmospheric biogeochemistry" allows for these changes in impacts depending on amount. For transport from the land to ocean, many constituents can also be carried by rivers; in the sections below, the relative importance of atmospheric versus riverine inputs is discussed. The atmosphere tends to be able to transport species more quickly and farther distances; however the mass of the atmosphere is much less that of water, providing a constraint on the amount of constituent the atmosphere can transport. For the cases of constituent transport from the oceans to land, or upstream on land, atmospheric transport on short timescales is often the most important, as the only other mechanism would be geological changes in orography.

The final important aspect of atmospheric biogeochemistry is that it allows us to understand the composition of the atmosphere, and the fluctuation in this composition, in an earth system context. Thus understanding atmospheric biogeochemistry represents an integrative approach to understanding the atmosphere and its relationship with the land, ocean, and crysophere.

Cycling in the Atmosphere

Since most nutrients cycle through the atmosphere, the atmospheric portion of their biogeochemistry is important. The most important nutrients to sustain life are carbon, nitrogen, and phosphorus, but many other nutrients are required. Here some of the most important nutrient cycles for the earth system are discussed, and cycles that are being perturbed due to human activities have to be focused upon.

Carbon Cycle

The carbon cycle receives substantial interest today because of the anthropogenic perturbation of the carbon cycle and the documentation of global warming. The carbon cycle represents the most important of the nutrient cycles, with the exception of water, for living organisms because living tissue is primarily composed of carbon. The dominant form of carbon in the atmosphere is carbon dioxide, a chemically inactive, but radiatively important species in the atmosphere. The riverine flux of carbon is 0.8 Tg/year, while the gross land and ocean fluxes with the atmosphere

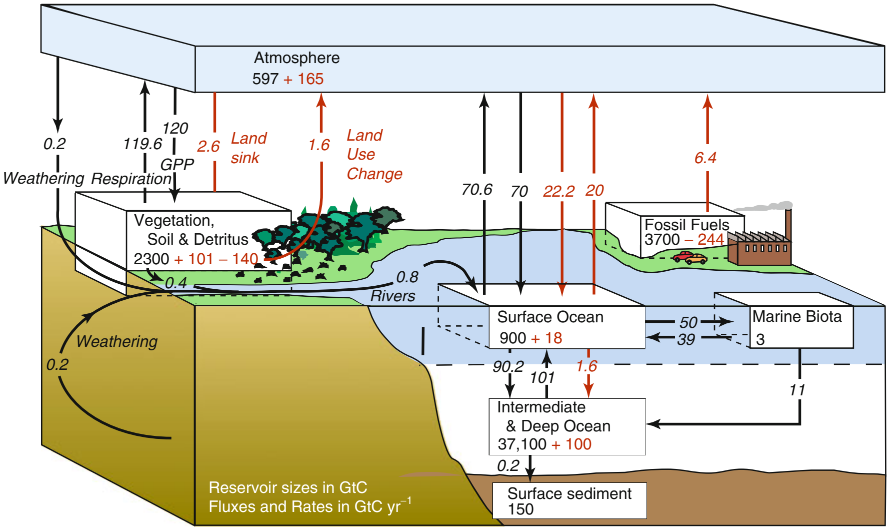


Fig. 2.1 The global carbon cycle for the 1990s, showing the main annual fluxes in PgC/year: preindustrial "natural" fluxes in black and "anthropogenic" fluxes in red. Gross fluxes generally have uncertainties of more than ±20% but fractional amounts have been retained to achieve overall balance when including estimates in fractions of PgC/year for riverine transport, weathering, deep ocean burial, etc. "GPP" is annual gross (terrestrial) primary production. Atmospheric carbon content and all cumulative fluxes since 1750 are as of end 1994 (Figure from [4])

are closer to 120 and 70 Tg/year (the net fluxes are closer to zero). This suggests that changes in the land or ocean carbon balance will be communicated rapidly through the atmosphere.

Carbon dioxide is removed from the atmosphere, not through chemical reactions, but through huge fluxes into the land or oceans, with an atmospheric lifetime of several years (Fig. 2.1 from [4]). In reality, however, the land and ocean fluxes are reasonably well balanced and it is difficult for all the extra carbon to be taken up, suggesting that the lifetime of the additional carbon dioxide emitted by humans is on the order of a few hundred years [4, 5]. However, there will be some carbon (perhaps 20–35%) that stays in the atmosphere for 3–7 kyr, thus the lifetime for atmospheric carbon dioxide is ambiguous [5].

Carbon dioxide shifts between glacial and interglacial climates are one of the most robust signals from the ice core records [6], and carbon dioxide tends to be about 80–100 ppm smaller during glacial periods than during interglacial periods (like today). These changes in carbon dioxide have not yet been explained, but are likely to be caused by a combination of colder temperatures, changes in ocean circulation, and changes in ocean productivity [7]. Indeed these changes in carbon dioxide are important forcing agents for maintaining the cold temperatures in the glacial periods [8], and thus must be an integral part of explaining the glacial/interglacial cycles over the past several hundred thousand years.

Currently, carbon dioxide emissions are increasing, and while preindustrial values of carbon dioxide were about 280 ppm, current values are over 380 ppm [4]. The reason for these changes in carbon dioxide values are due to human emissions of carbon dioxide release from combustion of fossil fuels and deforestation or other land-use change process. Carbon dioxide emissions currently (2008) are estimated to be about 8.7 +/− 0.5 PgC/year from fossil fuels, and another 1.5 +/− 0.7 PgC/year from land-use change (e.g., deforestation) [1]. Recent studies suggest that about 43% of anthropogenic carbon dioxide, from fossil fuel burning or deforestation, remains in the atmosphere, while the remainder is taken up by the land or the ocean, with more than half going into the ocean [1]. In the future, it is likely that less of the carbon dioxide that is emitted will be taken up by the land and ocean, due to both a saturation of the sinks of carbon dioxide, as well as the impact of climate change on the carbon cycle [2].

This increasing carbon dioxide in the atmosphere may increase land productivity and change land biogeochemistry. Higher carbon dioxide levels may allow plants to better adapt to water stress and can enhance productivity, a process called the carbon dioxide fertilization effect [9]. It is unclear how strong this fertilization effect is on land, because of limitation by other nutrients or water [10, 11]; however, available ecosystem studies support the hypothesis that increased carbon dioxide will result in higher productivity [9].

Higher carbon dioxide levels will tend to cause the ocean to take up more carbon, since the atmosphere will be supersaturated with respect to the oceans, especially in colder regions. This additional carbon dioxide is likely to lead to lower pH values in the oceans, called ocean acidification [12], and reduce the ability of calcifying organisms to create skeletons [13, 14]. This has potentially huge repercussions on ocean biogeochemistry and biodiversity [13, 15].

Carbon also exists in the atmosphere in other forms. The next most important constituent in terms of atmospheric mass is methane. Methane is formed by the decay of reduced organic species under anerobic conditions, occurring in wetlands, in geologic formations, or in the intestines of animals. Methane sources are estimated to be about 580 Tg/year, with anthropogenic sources (such as fossil fuel extraction or agriculture) representing 55–70% of the total emissions [4]. Methane is oxidized in the atmosphere by the OH radical with a lifetime of about 9 years, as well as smaller sinks of removal at the surface and photolysis in the stratosphere [4]. Methane is even more radiatively important than carbon dioxide on a per molecule basis, and thus its accumulation in the atmosphere is of concern for global warming. Methane concentrations have increased from ~650–1,770 ppb over the past couple hundred years, largely due to human activity [16]. Methane oxidation converts the methane into carbon dioxide.

Other important carbon-containing gases include carbon monoxide and non-methane hydrocarbons. Carbon monoxide is a product of incomplete combustion, and after carbon dioxide is the largest form of emission of carbon into the atmosphere (1,800–2,700 Tg CO/year) [17]. Carbon monoxide chemically reacts with the OH radical in the atmosphere, and has a lifetime of about 2 months [17]. The oxidation of carbon monoxide to form carbon dioxide represents an important sink of oxidants in the atmosphere [17]. Non-methane hydrocarbons refer to a large collection of carbon-containing compounds that are important for tropospheric photochemistry and the generation of ozone and oxidizing compounds. Non-methane hydrocarbons undergo oxidation via many mechanisms. The complete oxidation results in one mole of carbon dioxide for each carbon atom in the original molecule [17]. The lifetime of non-methane hydrocarbons vary considerably, but are usually less than the lifetime of methane. They are important for air quality, since they are precursors to ozone, and may themselves be toxic, but are not sufficiently abundant to have radiative impacts [17].

Carbon is also a component of atmospheric aerosols, but at much lower concentrations than the gases described above. These aerosols can, however, be important for radiation and chemistry, because they exist as a solid in a gas phase. Sources of carbonaceous aerosols include by-products from combustion processes, compounds emitted from vegetation, and in situ product from the condensation of non-methane hydrocarbons.

Nitrogen Cycle

The dominant form of nitrogen in the atmosphere is molecular nitrogen (N_2), which represents over 70% of the air. However, this form is chemically inert and not biologically available. Thus, the nitrogen cycle usually refers to the cycle of reactive nitrogen compounds. The dominant form of reactive nitrogen in the atmosphere is nitrous oxide, a long-lived gas (>70 years), which is destroyed predominately in the stratosphere by photolysis. The sources of nitrous oxide are in the soils, as by-products of nitrogen fixation and denitrification processes, which convert molecular nitrogen to reactive nitrogen and back. Humans have increased nitrous oxide from 270 to 319 ppb, largely due to an increase in nitrogen fertilizer application [4]. Emissions of nitrous oxide are estimated to be 11 Tg/year. Nitrous oxide is a potent green house, and represents the third largest radiative forcing gas (after carbon dioxide and methane) [4].

Potentially even more important than nitrous oxide are the nitrogen oxides (NO and NO_2), which currently have emissions of more than 140 Tg/year, while preindustrial emissions were thought to be closer to 30 Tg/year [18]. Nitrogen oxides are emitted during combustion process, with smaller contributions from soils and lightning. Nitrogen oxides have a short lifetime in the atmosphere (a few days), and are important for atmospheric photochemistry and air quality because of their role in producing ozone [17]. In addition, nitrogen oxides deposit onto the earth's surface as they are removed from the atmosphere, and thus move nitrogen from one region to another in large quantities (e.g., [19]).

Since preindustrial times, emissions and concentrations of ammonia have increased with current emissions of 52 Tg/year compared with an estimated 22 Tg/year in preindustrial times [18]. Ammonia is emitted predominately from natural soils and agriculture, with small contributions from oceans and energy use. Ammonia has a short lifetime in the atmosphere (a few days) and is removed through wet or dry deposition onto the surface.

Similar to carbon species, nitrogen is also contained in organic gas species and to a smaller extent in inorganic and organic aerosols. These species can act as reservoirs of reactive nitrogen (e.g., [17], and thus be important for atmospheric chemistry. The fraction of nitrogen species that are in aerosols is dependent on chemical equilibrium with other aerosols, especially the strong acid sulfate [20]. In the future, as sulfate emissions decrease, nitrogen aerosols may become more important in terms of the radiative forcing of the climate as the fraction of nitrogen in aerosol form increases [20].

Riverine inputs to oceans are thought to be 55–80 Tg/year [21], which is of the same magnitude as the fluxes of many nitrogen compounds discussed above, suggesting that the atmosphere plays an important role in the cycling of nitrogen between the land and ocean.

Because many terrestrial ecosystems are nitrogen limited, the impact of nitrogen deposition, especially from nitrogen oxides and ammonia, is thought to be large for land ecosystems [3, 22–25]. Nitrogen is also a limiting nutrient for ocean systems, and there have been large increases in nitrogen deposition to ocean systems [21]; however, because of the large reservoir of reactive nitrogen in the oceans, deposition to the oceans is not thought to substantially modify the ocean nitrogen budget [26].

Phosphorus Cycle

Unlike carbon or nitrogen, most of the phosphorus in the atmosphere is in the aerosol form, with little phosphorus in the gas phase (e.g., [27]. As phosphorus is approximately 700 ppm in crustal material, and mineral aerosols (soil particles suspended in the atmosphere) represent a major source of aerosols, most of the phosphorus in the atmosphere is in mineral aerosols [28] at about 1.15 TgP/year. The next largest source is thought to be primary biogenic particles (0.16 TgP/year), but these aerosols are very poorly understood [28]. Combustion sources from fossil fuels, biomass burning, and biofuels represent a small source (0.0.7 TgP/year), while volcanoes and sea salts are thought to provide small amounts as well [28]. Thus, continental regions, especially deserts are the dominant source of atmospheric phosphorus [28].

Aerosols deposited to ocean have a limited residence time in the ocean mixed layer, before they sink to the ocean floor and become incorporated into the sediment. Only the fraction of phosphorus which is soluble or bioavailable can contribute to the phosphorus cycle of the ocean [28, 29]. The bioavailable fraction is often considered phosphate or soluble-reactive phosphorus [28], and the fraction of phosphorus in aerosols that is soluble varies from 7% to 100% [28, 29]. Transects of the North Atlantic suggest that desert dust aerosols average about 10% solubility [30, 31]; however a gradient in solubility is observed as the particles move away from the desert dust source area, suggesting that atmospheric processing may make the phosphorus in aerosols more soluble [30]. Studies in Mediterranean suggest that acidification, partly due to anthropogenic emissions of sulfur and nitrogen species, can make the phosphorus in desert dust more soluble [32]. Non-dust phosphorus is thought to be much more soluble (∼50%), based on studies of European-derived aerosols in the Mediterranean [28, 33–36].

Phosphine gas (PH_3) has been measured in limited studies, with concentrations of between 0.39 and 2.45 ng/m^3 in remote regions and up to 178 ng/m^3 in urban locations or near paddy fields [37–39]. While traditionally, the gas phase transfer of phosphorus has been considered negligible [28, 40], some of the values listed here are similar in magnitude as those found for phosphorus in aerosols [28]. This suggests that phosphine could be an important mechanism for transferring phosphorus, yet is not well understood. Small amounts of phosphine could be generated in soils, agricultural and industrial processes, and lightning, but phosphine in the presence of sunlight is converted to phosphoric acid [37, 38].

Because phosphorus limitation is thought to be widespread in tropical forests and savannahs [41, 42], atmospheric deposition of desert dust may play a role in the long-term viability of tropical soils. For example, it has been suggested that the atmospheric deposition of desert dust from North Africa is responsible for the maintenance of the Amazon forest [43], and that deposition from Asia is important source of phosphorus to the tropical forests in Hawaii on long timescales [44]. The atmospheric deposition of phosphorus is likely to be important in many land ecosystems on geological timescales [45]. Forest and grassland ecosystems can also lose phosphorus through the atmosphere, as primary biogenic particles or biomass burning particles contain a large proportion of the phosphorus in tropical forests [46, 47].

The ocean tends to be a net sink of phosphorus from the atmosphere [28], and since productivity in the ocean is thought to be phosphorus limited on long timescales [48] and in some regions [49, 50], the atmospheric deposition of phosphorus could be an important source. Riverine inputs of phosphorus to the oceans are much larger than the atmospheric fluxes described here (11 Tg/year) [51]; however this phosphorus is input to the oceans in the coastal regions, and it is unclear how much of this phosphorus is advected into the open oceans [52]. However, on short timescales, large increases (25%) in phosphorus deposition to oceans which could be due to human activity [28] is not thought to substantially impact ocean biogeochemistry, because of the large reservoir of phosphorus already in the oceans [26].

Sulfur Cycle

Sulfur is an important trace compound in the earth system, used by some microorganisms for energy. It is commonly found in rocks with smaller quantities measured in the atmosphere [27]. The atmospheric portion of the sulfur cycle is important because of the climate interactions of the sulfur species, especially in the sulfate aerosol form. The sulfur compound with the longest lifetime (5 years) and most common by mass is carbonyl sulfide (OCS) [27]. The dominant sources (10 GgS/year) of carbonyl sulfide are the oceans and soils, with small contributions from combustion processes and the chemical oxidation of CS_2 [27]. Carbonyl sulfide is destroyed by oxidation in the atmosphere, as well as by stratospheric phytolysis and uptake by vegetation at the surface [27, 53]. Carbonyl sulfide is thought to contribute about half of the sulfur that forms the stratospheric aerosol layer, the other half coming from volcanic emissions [54].

Sulfur dioxide is a naturally occurring compound released from volcanoes and biomass burning ($<10\%$ total) as well as oxidation of the naturally emitted dimethyl sulfide ($\sim20\%$), but the dominant source currently are human combustion activities with a total emission of about 91–125 TgS/year [16, 18]. Once released into the atmosphere, sulfur dioxide is oxidized within a few days to form sulfate aerosols. Because sulfate is highly hydrophilic, sulfate aerosols take up water and make a larger contribution to aerosol-radiation interactions than their mass would suggest [16]. In addition these aerosols readily interact with cloud droplets to modify cloud optical properties (aerosol indirect effect on clouds) [16]. Sulfur dioxide from human emissions is responsible for a large portion of the anthropogenic radiation forcing of aerosols [16].

Dimethyl sulfide is a natural source of sulfur gas to the atmosphere from the ocean, with a magnitude of about 15 TgS/year, which is readily oxidized to sulfur dioxide and then to sulfate aerosol (lifetime about 1.5 days). This is the dominant gas phase species of sulfur released from the ocean to the atmosphere. Because of sulfate aerosol interactions with clouds, it was hypothesized that biota could modulate their temperature by modifying emissions of dimethyl sulfide [55]; however, recent studies suggest this mechanism is not a dominant mechanism, for example, under anthropogenic climate change [56].

Sea salts are 7.7% sulfate by weight [57], and represents the largest source of sulfur to the atmosphere. Approximately 10,000 Tg/year of sea salt aerosols [58] is emitted into the atmosphere, which translates to 770 Tg SO_4/year or 250 TgS/year. Sea salt aerosols reside in the atmosphere from hours to days, and tend to redeposit back onto oceans [58], and are important as cloud condensation nuclei [59].

Sulfur moves from the land to the ocean at a rate of approximately 130 TgS/year [27], which is similar in magnitude to many of the atmospheric fluxes, suggesting that atmospheric sulfur is an important part of the global biogeochemical cycle.

Iron Cycle

Iron is required in small quantities by almost all organisms (e.g., [48, 60]), and is often considered a micronutrient. Because some regions of the oceans are iron limited [61], atmospheric deposition of iron is important for providing new iron to ocean biota [62]. Riverine inputs of iron are large, but are largely removed in coastal regions [63]. There is some evidence that the nitrogen-fixing organisms have larger iron requirements

Similar to the phosphorus cycle, iron resides almost entirely in atmospheric aerosols, predominately in desert dust. Combustion and other sources of iron are

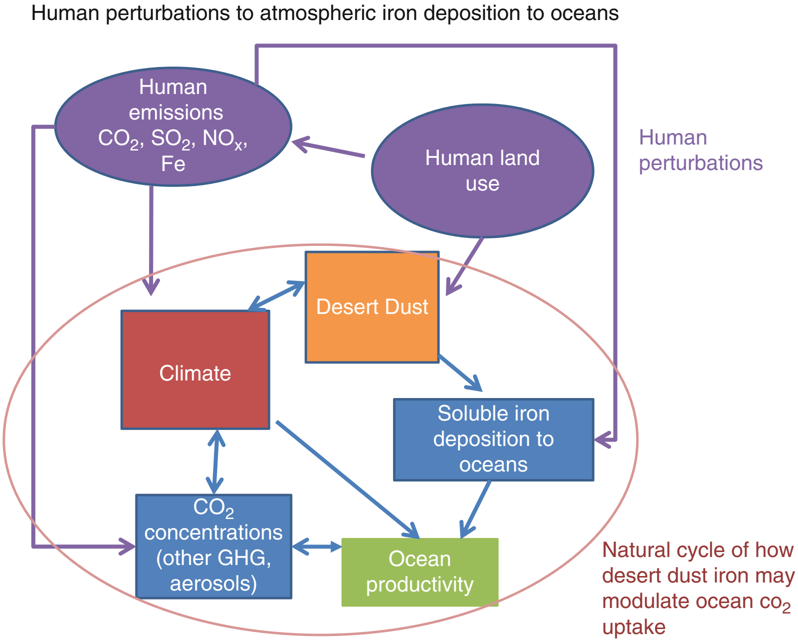


Fig. 2.2 Skematic representing feedbacks between natural ocean carbon cycle, carbon dioxide concentrations, and iron inputs also shows humans could be perturbing the iron deposition (Figure from [65])

estimated to be less than 5% of the total iron in aerosols [64, 65]. Crustal material is on average about 3.5% iron [66], with some minerals having substantially higher or lower concentrations [67, 68]. However, estimates and observations suggest that desert dust aerosols vary in their iron content by only a factor of 2, suggesting that the high heterogeneity in the soils is mixed in the atmosphere [65, 69]. There are also small sources of atmospheric iron from volcanoes [70, 71], cosmic dust [72], and combustion [64].

Because atmospherically deposited dust only resides in the mixed layer of the ocean for a short time, many researchers consider the soluble fraction of the iron the most relevant for ocean iota [73]. However, which fraction of the iron is really bioavailable is not well understood, but is likely to be a small fraction of the total iron in aerosols (1–80%) (see reviews in [65, 73]). Because the soluble fraction of iron in soils is much smaller than what is observed in the atmosphere, it is thought that atmospheric processing of iron is important [73, 74]. It is likely that the acidity and insolation play a role in processing of iron [73–75]. Some iron-containing minerals are more easily solubilized than others and iron in combustions is significantly more soluble than iron in mineral aerosols; these factors complicate our understanding of iron solubility [68, 76–78] (Fig. 2.2 from [65]).

Some regions of the ocean are iron deficient and additions of iron can result in phytoplankton blooms [79, 80]; however, there is limited evidence showing conclusively that a case of deposition of atmospheric iron results in a bloom [81, 82]. Additional atmospheric iron deposition could also lead to increases in ocean nitrogen, because of the potential for higher iron requirements for nitrogen-fixing organisms [48, 83]. There is some evidence for this effect [84–86], although some of these effects may be due to the isotopic signature of direct anthropogenic nitrogen deposition [87]. It is likely that human changes in iron deposition are more important than the large increase in direct nitrogen deposition to the ocean, in terms of both the nitrogen budget of the ocean, as well as carbon dioxide changes [26, 88].

Estimated changes in soluble iron due to increases in desert dust sources as well as the increase in atmospheric acidification (associated with more sulfate and nitrate aerosols) could be as large as a factor of 2–4 globally averaged [65, 89]. Model results suggest this would lead to an increase in the nitrogen fixation in the oceans, and a small sink of carbon dioxide [89, 90].

Oxygen Cycle

The atmospheric concentrations of oxygen are one of the clearest indications that life exists on this planet [27, 91]. The bulk of the atmospheric oxygen is in the gaseous state, in the form of molecular oxygen (O_2), which represents 21% of the atmosphere. Atmospheric oxygen is changing slightly due to human activity. For every molecule of CO_2 formed through combustion, one molecule of O_2 is lost from the atmosphere, and while small, these changes can be detected and provide important information about the carbon cycle [92, 93]. This molecular oxygen is also important as it absorbs high-energy uv light from the sun [94].

The oxygen species that exist in very small quantities in the atmosphere are extremely important for atmospheric composition, and control the oxidation of reduced species such as methane or nitrogen oxides. Ozone is produced in the stratosphere in the presence of high-energy uv light, and this ozone, as well as the process which generates the ozone, protects life from uv light at very short wavelengths [17, 94]. This also changes the thermal structure of the atmosphere, contributing to the increasing temperatures in the stratosphere [94]. Stratospheric ozone has been impacted by the release of chlorofluorocarbons, which are inert enough to reach the upper atmosphere, where they are photolyzed [95]. The largest impact has been in the polar regions, where an “ozone hole” has appeared during the hemispheric spring, although midlatitude reductions in ozone have also been observed [96]. In the troposphere, human emissions of nitrogen oxides, carbon monoxide, and non-methane hydrocarbons have led to an increase in ozone [17]. While the bulk of the ozone remains in the stratosphere, ozone is a highly oxidizing substance, so that increases in tropospheric ozone impacts human health and materials [17]. Ozone acts as a greenhouse gas, and the observed changes in ozone, both in the troposphere and stratosphere, have impacted the radiative budget [16].

Oxidation of organic and reduced species in the troposphere is the sink for many important greenhouse gases (e.g., methane), and other important trace species in the atmosphere (e.g., sulfur dioxides, nitrogen oxides) [17]. Much of this oxidation does not occur directly by ozone, but rather by the hydroxyl radical, a by-product of ozone photochemistry in the troposphere [17]. Because this constituent reacts so quickly, it is difficult to determine its concentration directly, but it can be deduced from changes in concentrations in some of the compounds it reacts with [97]. It is unclear whether humans are increasing or decreasing the hydroxyl radical, and will continue to do so, because of the complex nonlinear chemistry [16].

Other Cycles

Most elements are cycled in the atmosphere, and because long-range transport is much faster in the atmosphere than on land or oceans, this transport can be important for many elements. Here the focus is on a few critical elements for global biogeochemistry. However, some elements are important for their negative impacts, especially heavy metals. For example, mercury emissions from humans impacts human and ecosystem health [98]. Some studies have suggested that atmospheric deposition of heavy metals far from source areas can adversely impact ocean biogeochemistry [99]. In addition, atmospheric deposition of acid is thought to reduce the ability of some ecosystems to retain some nutrients and to negatively impact land and aquatic ecosystems [101–103], and to enhance ocean acidification [100].

Anthropogenic Modifications to Atmospheric Biogeochemical Cycles

Direct human emissions of carbon dioxide, sulfur dioxide, and nitrous oxides gases, as well carbonaceous aerosols, through the increase in combustion that humans use to generate energy, has substantially increased the sources and atmospheric concentrations of these compounds [16] (Fig. 2.3). In addition, changes in land use have increased carbon dioxide, methane, nitrous oxide, and ammonia emissions [16] (Fig. 2.3). These modifications are likely to continue into the future, with some decrease in emissions of combustion-sourced compounds estimated (Fig. 2.3) [101–107]. A combination of climate change and land-use change seems to have resulted in an increase in mineral aerosol over the past 100 years [89], which increases the flux of phosphorus and iron. In addition, bioavailable iron and phosphorus are likely to be even more susceptible to acidity and combustion sources; thus these compounds have changed more than simply from mineral aerosol changes over the past 100 years (Fig. 2.3). All of these factors have accelerated the atmospheric biogeochemical cycling of these compounds.

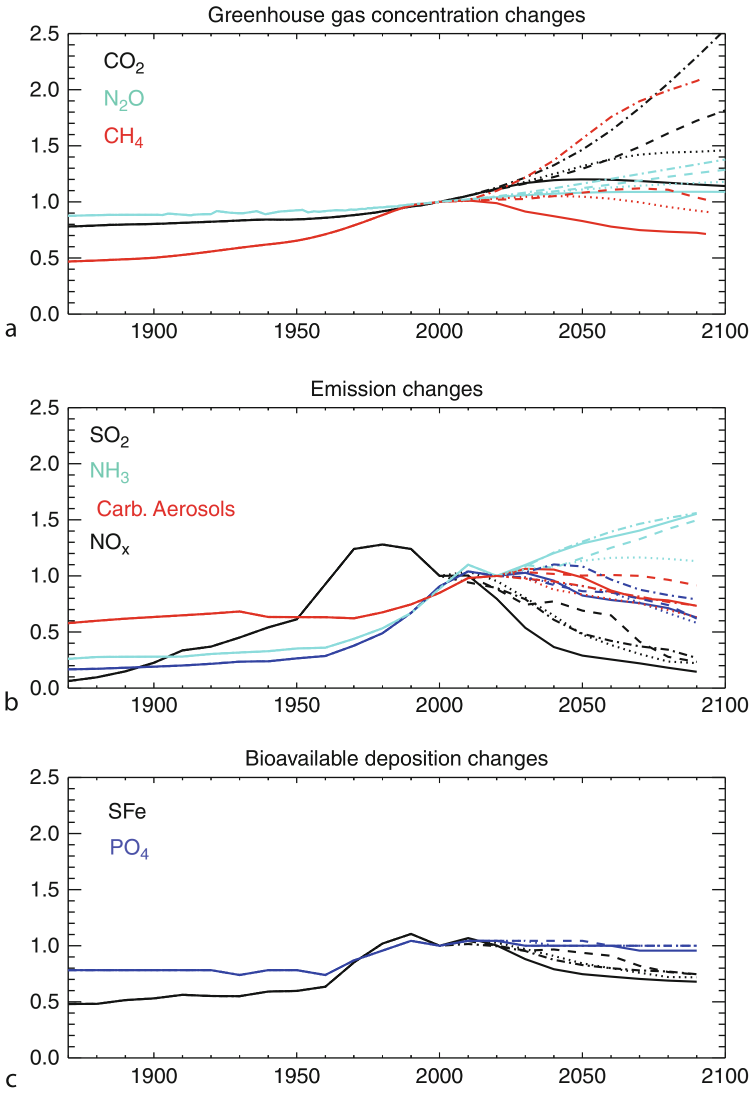


Fig. 2.3 Relative changes in green house gas concentrations (**a**), for carbon dioxide (*black*), nitrous oxide (*cyan*), and methane (*red*), relative to concentrations in 2000, based on atmospheric and ice core measurements for the past time period [127–130]. For the future, estimates are based on the representative concentration pathways estimates for four different scenarios [101–107], represented by the four lines into the future. Emission changes for important species, all normalized to 2000 values (**b**) for sulfur dioxide (*black*), ammonia (*cyan*), carbonaceous aerosols (*red*), and nitrogen oxides (*blue*) based on historical estimates [18], and for future estimates, the representative concentration pathways [101–107]. Deposition changes for bioavailable iron (*black*) and phosphorus (*blue*) are shown relatively to 2000 estimates (**c**), based on historical estimates and future estimates, using estimates of desert dust changes from [89], assuming in the future the model mean for no carbon dioxide fertilization changes in source area from [131], with iron solubility estimates from [64, 65], scaled by the sulfur dioxide emissions changes shown in (**b**), and for combustion soluble iron sources from [64] scaled by the black carbon emission changes included in (**b**). For soluble phosphorus (phosphates), desert dust estimate changes are the same as described for iron, and combustion sources of phosphorus come from [28], and are scaled by black carbon emissions, included in (**b**)

Atmospheric Modification to Other Biogeochemical Components

Atmospheric Biota

Atmospheric aerosols are likely to be composed in part by primary biogenic particles, or particles derived from the biota, for example, plant or insect bits, pollen, bacteria, or viruses [108]. Recent estimates suggest less than 5% of the aerosols are primary biogenic particles [109]. Some small fraction of these biogenic aerosols may contain viable living cells. While estimates of culturable bacteria are dependent on the methods used, estimates suggest that some bacteria do remain viable in the atmosphere, and can perhaps grow again if deposited onto the ground [110, 111]. Clearly, plants depend on pollen, and fungi on spores, to move genetic material from one place to another, and some fraction of this pollen and spores can be carried far from the source [112]. The amount of viable biota moving through the atmosphere and the importance of this process for global biogeochemical cycling is not well understood, but deserves future study.

Climate Modification of Biogeochemical Cycles

Changes in atmospheric biogeochemistry can modify climate, which in turn can also modify land and ocean biogeochemistry. This feedback has been used as a mechanism for understanding how carbon dioxide can be modulated on long timescales. For example, it is hypothesized that there are higher temperatures and more precipitation on land during higher carbon dioxide levels, thus increasing weathering [113]. This will reduce the atmospheric concentrations of carbon dioxide levels, acting as a feedback to prevent a runaway green house effect [113].

In the anthropocene, modeling studies have suggested that changes in anthropogenic aerosols and mineral aerosol particles can modify precipitation, incoming solar radiation, and temperature, and therefore modify land and ocean biogeochemistry enough to be seen at the global level [89, 90, 114, 115].

The Atmosphere as an Integrator

The atmosphere is the one part of the earth system that is transparent in many wavelengths of light, and thus can be readily observed using remote sensing methods. In addition, we live at the interface of the atmosphere and land surface, and thus have easy access to the lower layers of the atmosphere. This means that we have much better observations of the state of many biogeochemical cycles in the atmosphere, than in the land or ocean [16]. Thus understanding atmospheric biogeochemistry provides important information for understanding global biogeochemical cycles.

Transport in Atmosphere

The distance a chemical travels from its source region is highly dependent on its atmospheric lifetime. Many of the biogeochemically relevant gases discussed above are long-lived gases, which will reside in the atmosphere for several months to hundreds of years (e.g., CO_2, CH_4, N_2O). Other, reactive species (such as the hydroxyl radical OH), can have very short lifetimes (<1 day), and thus will not travel far from the sources. Aerosol species tend to have lifetimes between 1 day and a few weeks, and this lifetime is dependent on the size of the particle as well as whether precipitation has occurred, as this will remove particles.

For chemical species that have a relatively long lifetime (> a few days), long-range transport can occur. Once emitted into the atmosphere, constituents mix quickly into the atmospheric boundary layer (the lowest 1–3 km of the atmosphere, where mixing is strongest), and can be transported long distances. Air parcels stay on a constant buoyancy surface (called isentropes), unless they are heated or cooled [116]. These surfaces of constant buoyancy tilt upward between the tropics and the high latitudes, because the high latitudes are colder (Fig. 2.4a from [109]). An air parcel emitted in the tropical boundary layer can reach the stratosphere (>10 km high) in the high latitudes without being heated by moving along an isentropic surface. On average, most of the atmosphere is radiatively cooling slightly during non-stormy conditions, so, over time, parcels of air will become heavier and move downward. If the parcel encounters a storm system with clouds and precipitation, the parcel can experience large heating or cooling and move vertically very quickly, perhaps traveling from the boundary layer up to 10 km in less than 1 h [117]. At the same time, the precipitation within the storm system can cause wet deposition of soluble gases or aerosols. Based on atmospheric heating rates, the residence time for a parcel on a given isentropic surface varies between <1 day for the tropics to 3–6 days in midlatitudes (based on 75 percentile heating rates) (Fig. 2.4a). If a parcel moves away from the pole, the tilting of the neutral buoyant surfaces will direct the parcel downward vertically and vice versa.

Geographically, there tend to be particular regions in midlatitudes that have more storms (Fig. 2.4b), called the storm tracks. Winds in the midlatitudes (30–60°) tend to be from the west, while winds in the tropics and highlatitudes (>60°) are less uniform. Tropical aerosols are more likely to encounter storms, resulting in vertical mixing or quick removal (Fig. 2.4b). Transport across latitudes tends to occur less often than transport along a latitude [118]. Thus, in the midand high latitudes, one can think of a parcel containing a constituent moving along isentropes, following the local winds, slowly cooling, and moving to a lower isentropic surface until it encounters a storm system, at which time it can be quickly removed or vertically transported.

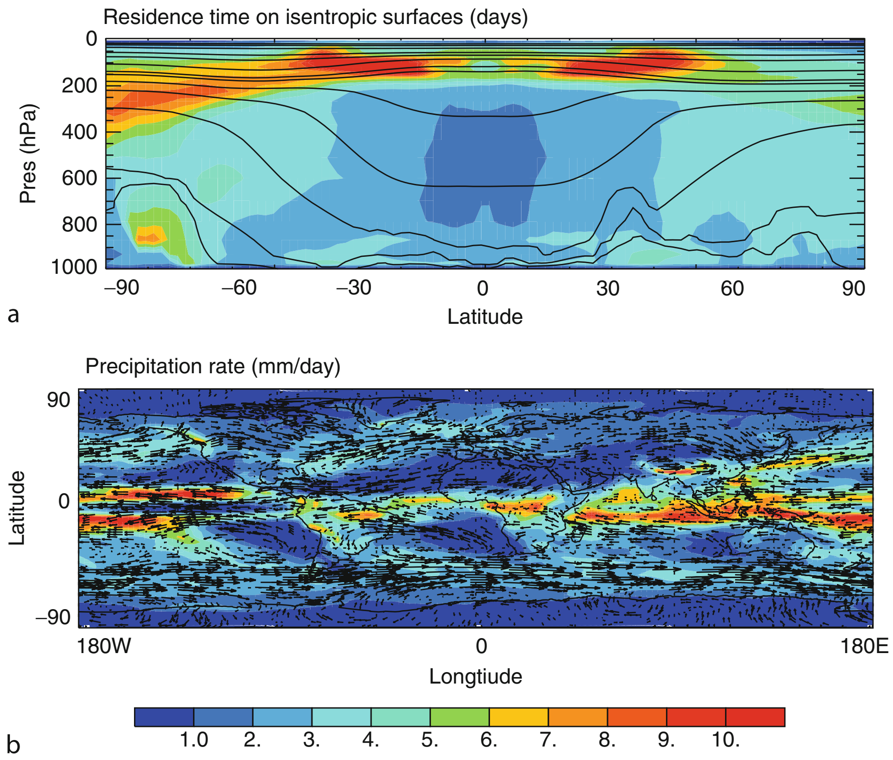


Fig. 2.4 Transport pathways of atmospheric constituents. Vertical distribution with latitude of the residence time of aerosols along a neutrally buoyant layer (isentrope), considering only vertical motion by heating in colors (**a**), with *black lines* marking surfaces of constant buoyancy (isentropes). These results are based on the 75% heating rates (warming or cooling) of the six hourly instantaneous results. Annually averaged precipitation rate (mm/day) in color, with annual mean wind vectors for 850 hPa (just above the boundary layer) (**b**) (Figure from [109])

Inverse Modeling

Inverse modeling is the practice of using atmospheric concentrations to deduce the sources and sinks of various important compounds [119]. Because many biogeochemically relevant species have natural sources and sinks, they have sources of uncertain magnitude. Inverse modeling has become quite common for these constituents. The initial studies used simple models to best match available data [120, 121]. With time, studies focused on regional budgets but used higher temporal and spatially resolved datasets and models to resolve regional and time-varying sources as well as sophisticated statistical techniques [122–124]. These studies (called top-down studies, because they use atmospheric concentrations to constrain sources and sinks) provide important information about the sources and can provide alternative views from upscaling flux measurements or inventories (called bottom-up approaches) [125]. Current approaches often estimate parameters based on available data, instead of resolving budgets, allowing models to be tuned to best match observations at a site, before applying that model to other regions to extrapolate to regional and global budgets [126]. Thus understanding atmospheric biogeochemistry of important compounds can facilitate a better understanding of the land and ocean biogeochemistry, using inverse modeling approaches.

Future Directions

The study of atmospheric biogeochemistry as a field is just beginning, although some of the most important elements, such as the carbon cycle, have been studied for many years. Our knowledge, however, of even the carbon cycle is insufficient to understand how global carbon dioxide levels will respond in the future, even assuming that human emissions can be predicted [2]. Important nutrient cycles highlighted here include the carbon, nitrogen, phosphorus, sulfur, iron, and oxygen cycles. All of these cycles in some way are heavily impacted by human activities, and in addition, feedback onto our climate and environment in important ways. Many of the important innovations over the last few years have involved understanding how the different cycles interact. Thus atmospheric biogeochemistry remains an important field for continued research and innovations.

The atmosphere is unique with respect to the other components of the earth system in our ability to easily observe many layers using remote sensing. This means that the atmosphere provides a location where biogeochemistry as a whole can be better understood. Inverse methods and modeling, in addition to the many observations, may allow us to improve our understanding of not just atmospheric biogeochemical cycles, but land and ocean biogeochemical cycles.

Acknowledgments The author would like to thank NSF (0932946, 0832782, 0758369) and NASA (NNG06G127G), as well as Rachel Scanza for assistance on the manuscript.

Bibliography

Primary Literature

1. LeQuere C, Raupach M, Canadell J, Marland G, Bopp L et al (2009) Trends in the sources and sinks of carbon dioxide. Nat Geosci 2:831–836. doi:10.1038/ngeo689
2. Friedlingstein P, Cox P, Betts R, Bopp L, Von Bloh W et al (2006) Climate-carbon cycle feedback analysis, results from the C4MIP Model intercomparison. J Clim 19:3337–3353
3. Thornton P, Doney S, Lindsay K, Moore JK, Mahowald N et al (2009) Carbon-nitrogen interactions regular climate-carbon cycle feedbacks: results from an atmosphere–ocean general circulation model. Biogeosciences Discussion 6:3303–3354

4. Denman KL, Brasseur G, Chidthaisong A, Ciais P, Cox PM et al (2007) Couplings between changes in the climate system and biogeochemistry. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science bases. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK
5. Archer D, Eby M, Brovkin V, Ridgwell A, Cao L et al (2009) Atmospheric lifetime of fossil fuel carbon dioxide. Annu Rev Earth Pl Sc 37:117–134
6. Petit JR, Jouzel J, Raynaud D et al (1999) Climate and atmospheric history of the past 420,000 years from the Vostok Ice core, Antarctica. Nature 399:429–436
7. Sigman DM, Boyle EA (2000) Glacial/interglacial variations in atmospheric carbon dioxide. Nature 407:859–869
8. Hewitt C, Mitchell J (1997) Radiative forcing and response of a GCM to ice age boundary conditions: cloud feedback and climate sensitivity. Clim Dynam 13:821–834
9. Long SP, Ainsworth E, Rogers A, Ort D (2004) Rising atmospheric carbon dioxide: plants FACE the future. Annu Rev Plant Biol 55:591–628
10. Caspersen J, Pacala S, Jenkins J, Hurtt G, Moorcroft P, Birdsey R (2000) Contributions of land-use history to carbon accumulation in U.S. forests. Science 290:1148–1151
11. Joos F, Prentice IC, House JI (2002) Growth enhancement due to global atmospheric change as predicted by terrestrial ecosystem models: consistent with US forest inventory data. Global Change Biol 8:299–303
12. Caldeira K, Wickett M (2003) Anthropgenic carbon and ocean pH. Nature 425:365
13. Kleypas J, Buddemeier R, Archer D, Gattuso J-P, Langdon C, Opdyke B (1999) Geochemical consequences of increased atmospheric carbon dioxide on coral reefs. Science 284:118–120
14. Riebesell U, Zondervan I, Rost B, Tortell P, Zeebe R, Morel F (2000) Reduced calcification of marine plankton in respnse to increased atmospheric CO_2. Nature 407:364–367
15. Doney S, Fabry V, Feely R, Kleypas J (2009) Ocean acidification: the other CO_2 problem. Annu Rev Mar Sci 1:169–192
16. Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R et al (2007) Changes in atmospheric constituents and in radiative forcing. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK/New York, pp 130–234
17. Jacob D (1999) Atmospheric chemistry. Princeton University Press, Princeton, p 266
18. Lamarque J-F, Bond T, Eyring V, Granier C, Heil A et al (2010) Historical (1850–200) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. Atmos Chem Phys 10:7017–7039
19. Galloway J, Townsend A, Erisman J, Bekunda M, Cai Z et al (2008) Transformation of the nitrogen cycle: recent trends, questions and potential solutions. Science 320:889–892
20. Adams P, Seinfeld J, Koch D (1999) Global concentrations of tropospheric sulfate, nitrate and ammonium aerosol simulated in a general circulation model. J Geophys Res 104:13791–13823
21. Duce RA et al (2008) Impacts of atmospheric anthropogenic nitrogen on the open ocean. Science 320. doi:10.1126/science.1150369
22. Holland E et al (1997) Variations in the predicted spatial distribution of atmospheric nitrogen deposition and their impact on carbon uptake by terrestrial ecosystems. J Geophys Res 102:15849–15866
23. Townsend A, Braswell BH, Holland E, Penner J (1996) Spatial and temporal patterns in potential carbon storage due to deposition of fossil fuel derived nitrogen. Ecol Appl 6. doi:10.2307/2269486
24. Sokolov AP, Kicklighter DW, Melillo JM, Felzer BS, Schlosser CA, Cronin T (2008) Consequences of considering carbon-nitrogen interactions on the feedbacks between climate and the terrestrial carbon cycle. J Clim 21:3776–3796
25. Zaehle S, Friedlingstein P, Friend AD (2010) Terrestrial nitrogen feedbacks may accelerate future climate change. Geophys Res Lett 37. doi:10.1029/2009GL041345

26. Krishnamurthy A, Moore JK, Mahowald N, Luo C, Zender CS (2010) The impacts of atmospheric nutrient inputs on marine biogeochemistry. J Geophys Res 115. doi:10.1029/2009JG001115
27. Schlesinger WH (1997) Biogeochemistry: an analysis of global change. Academic, San Diego, 588 pp
28. Mahowald N, Jickells TD, Baker AR, Artaxo P, Benitez-Nelson CR, Bergametti G, Bond TC, Chen Y, Cohen DD, Herut B, Kubilay N, Losno R, Luo C, Maenhaut W, McGee KA, Okin GS, Siefert RL, Tsukuda S (2008) The global distribution of atmospheric phosphorus deposition and anthropogenic impacts. Global Biogeochem Cycle 22. doi:10.1029/2008GB003240
29. Graham WF, Duce RA (1982) The atmospheric transport of phosphorus to the Western North Atlantic. Atmos Environ 16:1089–1097
30. Baker A, French M, Linge K (2006) Trends in aerosol nutrient solubility along a west–east transect of the Saharan dust plume. Geophys Res Lett 33. doi:10.1029/2005GL024764
31. Baker A, Jickells T, Biswas K, Weston K, French M (2006) Nutrients in atmospheric aerosol particles along the Atlantic Meridional Transect. Deep-Sea Res II 53:1706–1719
32. Nenes A, Krom M, Mihalopoulos N, VanCapellen P, Shin Z et al (2010) Atmospheric acidification of mineral aerosols: a source of bioavailable phosphorus to the oceans. Atmos Chem Phys Discuss 11:6163–6185
33. Bergametti G, Remoudaki E, Losno R, Stiner E, Chatenet B, Buat-Menard P (1992) Source, transport and deposition of atmospheric phosphorus over the Northwestern Mediterranean. J Atmos Chem 14:501–513
34. Herut B, Krom M, Pan G, Mortimer R (1999) Atmospheric input of nitrogen and phosphorus to the Southeast Mediterranean: sources, fluxes and possible impact. Limnol Oceanogr 44:1683–1692
35. Herut B, Collier R, Krom M (2002) The role of dust in supplying nitrogen and phosphorus to the Southeast Mediterranean. Limnol Oceanogr 47:870–878
36. Ridame C, Guieu C (2002) Saharan input of phosphate to the oligotrophic water of the open western Mediterranean Sea. Limnol Oceanogr 47:856–869
37. Glindemann D, Edwards M, Kuschk P (2003) Phosphine gas in the upper troposphere. Atmos Environ 37:2429–2433
38. Glindemann D, Edwards M, J-a L, Kuschk P (2005) Phosphine in soils, sludges, biogases and atmospheric implications–a review. Ecol Eng 24:457–463
39. Zhu R, Glindemann D, Kong D, Sun L, Geng J, Wang X (2007) Phosphine in the marine atmosphere along a hemisphere course from China to Antarctica. Atmos Env 41:1567–1573
40. Graham WF, Duce RA (1979) Atmospheric pathways of the phosphorus cycle. Geochimica et Cosmochimica Acta 43:1195–1208
41. Vitousek P (1984) Litterfall, nutrient cycling and nutrient limitations in tropical forests. Ecology 65:285–298
42. Okin GS, Mladenov N, Wang L, Cassel D, Caylor KK et al (2008) Spatial pattern of soil nutrients in two southern African savannas. J Geophys Res 111. doi:10.1029/2007JG000584
43. Swap R, Garstang M, Greco S, Talbot R, Kallberg P (1992) Saharan dust in the Amazon Basin. Tellus 44B:133–149
44. Chadwick OA, Derry LA, Vitousek PM, Huebert BJ, Hedin LO (1999) Changing sources of nutrients during four million years of ecosystem development. Nature 397:491–496
45. Okin G, Mahowald N, Chadwick O, Artaxo P (2004) The impact of desert dust on the biogeochemistry of phosphorus in terrestrial ecosystems. Global Biogeochem Cycle 18: GB2005. doi:10.1029/2003GB002145
46. Graham B, Guyon P, Maenhaut W, Taylor PE, Ebert M et al (2003) Composition and diurnal variability of the natural Amazonian aerosol. J Geophys Res 108:4765. doi:10.1029/2003JD004049
47. Mahowald N, Artaxo P, Baker A, Jickells T, Okin G et al (2005) Impact of biomass burning emissions and land use change on Amazonian atmospheric cycling and deposition of phosphorus. Global Biogeochem Cycle 19:GB4030. doi:10.1029/2005GB002541

48. Falkowski PG, Barber RT, Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. Science 281:200–206
49. Wu J, Sunda W, Boyle E, Karl D (2000) Phosphate depletion in the western North Atlantic Ocean. Science 289:759–762
50. Krishnamurthy A, Moore K, Zender C, Luo C (2007) Effects of atmospheric inorganic nitrogen deposition on ocean biogeochemistry. J Geophys Res 112. doi:10.1029/2006JG000334
51. Seitzinger S, Harrison J, Dumont E, Beusen A, Bouwman A (2005) Sources and delivery of carbon, nitrogen and phosphorus to the coastal zone: an overview of the Global Nutrient Export from Watersheds (NEWS) models and their application. Global Biogeochem Cycle 19. doi:10.1029/2005GB002606
52. Benitez-Nelson C (2000) The biogeochemical cycling of phosphorus in marine systems. Earh-Sci Rev 51:109–135
53. Crutzen PJ (1976) The possible importance of CSO for the sulfate layer of the stratosphere. Geophys Res Lett 3:73–76
54. Deshler T (2007) Stratospheric aerosol: measurements, importance, life cycle, anomalous aerosol. In: O'Dowd C, Wagner P (eds) Nucleation and atmospheric aerosols. Springer, Heidelberg, pp 613–624
55. Charlson RJ, Lovelock JE, Andreae M, Warren SG (1987) Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. Nature 326:655–661
56. Kloster S, Feichter J, Maier-Reimer E, Roeckner E, Stier P et al (2007) Response of DMS in the ocean and atmosphere to global warming. J Geophys Res 142:G03005. doi:10.1029/2006JG000224
57. Seinfeld J, Pandis S (1998) Atmospheric chemistry and physics. Wiley, New York, p 1326
58. Gong SL, Barrie L, Lazare M (2002) Canadian aerosol model (CAM): a size-segregated simulation of atmospheric aerosol process for climate and air quality models 2. Global sea-salt aerosol and its budgets. J Geophys Res 107:4779. doi:10.209/2001JD002004
59. Carslaw KS, Boucher O, Spracklen D, Mann G, Rae JG et al (2010) A review of natural aerosol interactions and feedbacks within the earth system. Atmos Chem Phys 10:1701–1737
60. Morell F, Price N (2003) The biogeochemical cycles of trace metals in the oceans. Science 300:944–947
61. Martin J, Gordon RM, Fitzwater SE (1991) The case for iron. Limnol Oceanogr 36:1793–1802
62. Fung I, Meyn SK, Tegen I, Doney S, John J, Bishop J (2000) Iron supply and demand in the upper ocean. Global Biogeochem Cycle 14:281–295
63. Poulton S, Raiswell R (2002) The low-temperature geochemical cycle of iron: from continental fluxes to marine sediment deposition. Am J Sci 302:774–805
64. Luo C, Mahowald N, Bond T, Chuang PY, Artaxo P et al (2008) Combustion iron distribution and deposition. Global Biogeochem Cycle 22. doi:10.1029/2007GB002964
65. Mahowald N, Engelstaedter S, Luo C, Sealy A, Artaxo P et al (2009) Atmospheric iron deposition: global distribution, variability and human perturbations. Annu Rev Mar Sci 1:245–278. doi:10.1146/annurev/marine.010908.163727
66. Duce RA, Liss PS, Merrill JT, Atlas EL, Buat-Menard P et al (1991) The atmospheric input of trace species to the world ocean. Global Biogeochem Cycle 5:193–259
67. Claquin T, Schulz M, Balkanski Y (1999) Modeling the mineralogy of atmospheric dust sources. J Geophys Res 104:22243–22256
68. Journet E, Desbouefs K, Caqineau S, Colin J-L (2008) Mineralogy as a critical factor of dust iron solubility. Geophys Res Lett 35. doi:10.1029/2007GL031589
69. Hand J, Mahowald N, Chen Y, Siefert R, Luo C et al (2004) Estimates of soluble iron from observations and a global mineral aerosol model: biogeochemical implications. J Geophys Res 109:D17205. doi:10.1029/2004JD004574
70. Frogner P, Gislason SR, Oskarsson N (2001) Fertilizing potential of volcanic ash in ocean surface water. Geol Soc Am 29:487–490

71. Duggen S, Croot P, Schacht U, Hoffmann L (2007) Subduction zo006Ee volcanic ash can fertilize the surface ocean and stimulate phytoplankton growth: evidence from biogeochemical experiments and satellite data. Geophys Res Lett 34:L01612. doi:10.1029/2006GL027522
72. Johnson K (2001) Iron supply and demand in the upper ocean: is extraterrestrial dust a significant source of bioavailable iron? Global Biogeochem Cycle 15:61–63
73. Jickells T, Spokes L (2001) Atmospheric iron inputs to the oceans. In: Turner DR, Hunteger K (eds) Biogeochemistry of iron in seawater. Wiley, Chichester, pp 85–121
74. Zhu X, Prospero J, Millero F (1997) Diel variability of soluble Fe(II) and soluble total Fe in North Africa dust int he trade winds at Barbados. J Geophys Res 102:21297–21305
75. Meskhidze N, Chameides W, Nenes A (2005) Dust and pollution: a recipe for enhanced ocean fertilization? J Geophys Res 110. doi:10.1029/2004JD005082
76. Guieu C, Bonnet S, Wagener T, Loye-Pilot M-D (2005) Biomass burning as a source of dissolved iron to the open ocean? Geophys Res Lett 22:L19608. doi:10.1029/2005GL022962
77. Chuang P, Duvall R, Shafer M, Schauer J (2005) The origin of water soluble particulate iron in the Asian atmospheric outflow. Geophys Res Lett 32. doi:10.1029/2004GL021946
78. Sedwick P, Sholkovitz E, Church T (2007) Impact of anthropogenic combustion emissions on the fractional solubility of aerosol iron: evidence from the Sargasso Sea. Geochem Geophy Geosy 8. doi:10.1029/2007GC001586
79. Coale KH, Johnson KS, Fitzwater SE, Gordon RM, Tanner S et al (1996) A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. Nature 383:495–501
80. Boyd PW, Law CS (2001) The southern ocean iron release experiment (SOIREE) introduction and summary. Deep-Sea Res II 48:2425–2438
81. Bishop J, Davis R, Sherman J (2002) Robotic observations of dust storm enhancement of carbon biomass in the north pacific. Science 298:817–821
82. Cassar N, Bender M, Barnett B, Fan S, Moxim WJ et al (2007) The southern ocean biological response to aeolian iron deposition. Science 317:1067–1070
83. Berman-Frank I, Cullen JT, Shaked Y, Sherrell RM, Falkowski P (2001) Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. Limnol Oceanogr 46:1249–1260
84. Michaels AF, Olson D, Sarmiento JL, Ammerman JW, Fanning K et al (1996) Inputs, losses and transformations of nitrogen and phosphorous in the pelagic North Atlantic Ocean. Biogeochemistry 35:181–226
85. Capone DG, Zehr JP, Paerl HW, Bergman B, Carpenter EJ (1997) Trichodesmium, a globally significant marine cyanobacterium. Science 276:1221–1229
86. Mahaffrey C, Williams R, Wolff G, Mahowald N, Anderson W (2003) Isotopic signals of nitrogen fixation over the eastern North Atlantic. Geophys Res Lett 40. doi:10.109/2002/GL016542
87. Knapp A, Hastings M, Sigman D, Lipschultz F, Galloway J (2010) The flux and isotopic composition of reduced and total nitrogen in Bermuda rain. Mar Chem 120:83–89
88. Okin G, Baker A, Tegen I, Mahowald N, Dentener F et al (2011) Impacts of atmospheric nutrient deposition on marine productivity: roles of nitrogen, phosphorus and iron. Global Biogeochem Cycle 25:GB2022
89. Mahowald N, Kloster S, Engelstaedter S, Moore JK, Mukhopadhyay S et al (2010) Observed 20th century desert dust variability: impact on climate and biogeochemistry. Atmos Chem Phys 10:10875–10893
90. Mahowald NM, Lindsay K, Rothenberg D, Doney SC, Moore JK et al (2011) Desert dust and anthropogenic aerosol interactions in the Community Climate System Model coupled-carbon-climate model. Biogeosciences 8:387–414. doi:10.5194/bg-8-387-2011
91. Lovelock J (1979) Gaia: a new look at life on earth. Oxford University Press, Oxford, UK
92. Keeling R, Shertz SR (1992) Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle. Nature 358:723–727

93. Bender M, Ho D, Hendricks M, Mika R, Battle M et al (2005) Atmospheric O_2/N_2 changes, 1993–2002: implications for the partitioning of fossil fuel CO_2 sequestration. Global Biogeochem Cycle 19. doi:10.1029/2004GB002410
94. Hartmann D (1994) Global physical climatology. Academic, San Diego
95. Solomon S (1999) Stratospheric ozone depletion: a review of concepts and history. Rev Geophys 37:275–316
96. Chipperfield MP, Randel WJ, Bodeker GE, Dameris M, Fioletov VE, AyitÈ-LÙ DLA, Ajavon N, MÈgie G, Watson RT et al (2003) Global ozone: past and future, chapter 4. In: Scientific assessment of ozone depletion: 2002. WMO, Geneva, p 498
97. Prinn RG, Weiss RG, Miller BR, Huang J, Alyea FN et al (1995) Atmospheric trends and lifetime of CH_3CCl_3 and global OH concentrations. Science 269:187–192
98. Selin N (2009) Global biogeochemical cycling of Mercury: a review. Annu Rev Env Resour 34:43–63
99. Paytan A, Mackey K, Chen Y, Lima I, Doney S et al (2009) Toxicity of atmospheric aerosols on marine phytoplankton. Proc Natl Acad Sci USA 106:4601–4605. doi:10.1073/pnas.0811486106
100. Doney S, Mahowald N, Lima I, Feeley R, Mackenzie F et al (2007) Impact of anthropogenic atmospheric nitrogen and sulfur depositionon ocean acidification and the inorganic carbon system. Proc Natl Acad Sci USA 104:14580–14585. doi:10.1073/pnas.07022181
101. vanVuuren D, den Elzen MG, Lucas P, Eickhout B, Strengers B et al (2007) Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. Clim Change 81:119–159. doi:10.1007/s10584-006-9172-9
102. Clarke L, Edmonds J, Jacoby H, Pitcher H, Reilly J, Richells R (2007) Scenarios of greenhouse gas emissions and atmospheric concentrations. Sub-report 2.1A of synthesis and assessment product 2.1 by the US climate change science program and the subcommittee on global change research. Department of Energy, Office of Biological and Environmental Research, Washington, DC
103. Smith SJ, Wigley T (2006) Multi-gas forcing stabilization with the MiniCAM. Energy J Special Issue #3:373–391
104. Wise M, Calvin KV, Thomson A, Clarke LE, Bond-Lamberty B et al (2009) Implications of limiting CO_2 concentrations for land use and energy. Science 324:1183–1186
105. Fujino J, Nair R, Kainuma M, Masui T, Matuoka Y (2006) Multi-gas mitigation analysis on stabilzation scenarios using AIM global model. Energy J Special Issue 3:343–354
106. Hijoka Y, Matuoka Y, Hisimoto H, Masui M, Kainuma M (2008) Global GHG emission scenarios under GHG concentration stabilization targets. Global Env Eng 13:97–108
107. Riahl K, Gruebler A, Nakicenovic N (2007) Scenarios of long-term socio-economic and environmental development under climate stabilization. Technol Forecast Soc 74:887–935
108. Jaenicke R (2005) Abundance of cellular material and proteins in the atmosphere. Science 308:73
109. Mahowald N, Ward D, Kloster S, Flanner M, Heald C et al (2011) Aerosol impacts on climate and biogeochemistry. Annu Rev Env Resour 36. doi:10.1146/annurev-environ-042009-094507
110. Lighthart B (2006) The ecology of bacteria in the alfresco atmosphere. FEMS Microbiol Ecol 23:263–274
111. Burrows SM, Elbert W, Lawrence MG, Poschl U (2009) Bacteria in the global atmosphere–Part 1: review and synthesis of literature for different ecosystems. Atmos Chem Phys 9:9263–9280
112. Heald C, Spracklen D (2009) Atmospheric budget of primary biological erosol particles from fungal sources. Geophys Res Lett 36. doi:10.1029/2009GL037493
113. Stanley S (1999) Earth system history. WH Freeman, New York, p 615
114. Jones C, Cox P, Essery R, Roberts D, Woodage M (2003) Strong carbon cycle feedbacks in a climate model with interactive CO_2 and sulphate aerosols. Geophys Res Lett 30:1479. doi:10.029/2003gl0166867

115. Mercado L, Bellouin N, Stich S, Boucher O, Huntingford C et al (2009) Impacts of changes in diffuse radiation on the global land carbon sink. Nature 458:1014–1018. doi:10.1038/nature07949
116. Holton JR (1989) Global transport processes in the atmosphere. In: Hutzinger O (ed) The handbook of environmental chemistry. Springer, New York, pp 97–144
117. Barth M, Stuart A, Skamarock W (2001) Numerical simulations of the July 10, 1996, stratospheric-tropospheric experiment: radiation, aerosols, and ozone (STERAO)-deep convection experiment storm: redistribution of soluble tracers. J Geophys Res 106:12381–12400
118. Bowman K, Carrie G (2002) The mean-meridional transport circulation of the troposphere in an idealized GCM. J Atmos Sci 59:1502–1514
119. Kasibhatla P, Heimann M, Rayner P, Mahowald N, Prinn R, Hartley D (eds) (2000) Inverse methods in global biogeochemical cycles. American Geophysical Union, Washington, DC, p 324
120. Bolin B, Keeling C (1963) Large-scale atmospheric mixing as deduced from teh seasonal and meridionial variations of carbon dioxide. J Geophys Res 68:3899–3920
121. Fung I, John J, Lerner J, Matthews E, Prather M et al (1991) Three-dimensional model syntehsis of the global methane cycle. J Geophys Res 96:13033–13065
122. Gurney KR, Law RM, Denning AS, Rayner PJ, Baker D et al (2002) Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. Nature 415:626–630
123. Bousquet P, Peylin P, Ciais P, Quere CL, Friedlingstein P, Tans PP (2000) Regional changes in carbon dioxide fluxes of land and oceans since 1980. Science 290:1342–1347
124. Houweling S, Kaminski T, Dentener F, Lelieveld J, Heimann M (1999) Inverse modeling of methane sources and sinks using the adjoint of ta global transprot model. J Geophys Res 104:26137–26160
125. Kasibhatla P, Arellano A, Logan J, Palmer P (2002) Top-down estimate of a large source of atmospheric carbon monoxide associated with fuel combustion in Asia. Geophys Res Lett 29. doi:10.1029/2002GL015581
126. Braswell B, Sacks W, Linder E, Schimel D (2005) Estimating diurnal to annual ecosystem parameters by synthesis of carbon flux model with eddy covariance net ecosystem exchange observations. Glob Change Biol 11. doi:1111/j.365-2486.005.00897
127. Keeling RF, Piper SC, Bollenbacher AF, Walker JS (2009) Atmospheric CO_2 records from sites in the SIO air sampling network. In: Trends: a compendium of data on global change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Dept. of Energy, Oak Ridge
128. Etheridge DM, Steele LP, Langenfelds RL, Francey RJ, Barnola J-M, Morgan VI (1998) Historical CO_2 records from the law dome DE08, DE08-2, and DSS ice cores. In: Trends: a compendium of data on global change. Oak Ridge National Laboratory: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge
129. Etheridge DM, Steele LP, Francey RJ, Langenfelds RL (2002) Historical CH_4 records since about 1000 A.D. from ice core data. In: Trends: a compendium of data on global change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge
130. Battle M, Bender M, Sowers T, Tans PP, Butler JH et al (1996) Atmospheric gas concentrations over the past century measured in air from firn at the South Pole. Nature 383:231–235
131. Mahowald N (2007) Anthropocence changes in desert area: sensitivity to climate model predictions. Geophys Res Lett 34. doi:10.1029/2007GL030472

Books and Reviews

Jacobson M, Charlson R, Rodhe H, Orians G (eds) (2000) Earth system science, vol 72. Academic, San Diego

Chapter 3
Ecological Succession and Community Dynamics

Herman H. (Hank) Shugart

Glossary

Abstract community	A group of organisms that recurs on the landscape with an implication of a level of integration among its parts that in extreme could be called organismal or quasi-organismal (see: Concrete Community).
Chronosequence	A space-for-time substitution in which the successional vegetation is ordered in a regular fashion. An example would be the series of vegetation at the foot of a receding glacier or a series of sand dunes ordered in regularly aged series.
Clementsian succession	An explanation of succession emphasizing the attributes of the community as if it functioned like a single living organism (see: Abstract Community).
Climax community	An ecological community associated with a particular climate and in a state of dynamic equilibrium with the climate. The American ecologist, F.E. Clements, designated the climax community as the endpoint of ecological succession for a given climate condition and noted the common Greek root κλίμα (*clima*) or inclination, in both climax and climate as indicating their close relationship.
Concrete community	The collection of organisms found at a specific place and time (see: Abstract Community).

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

H.H. (Hank) Shugart (✉)
Department of Environmental Sciences, The University of Virginia, 376 Clark Hall,
Charlottesville, VA 22904-4123, USA
e-mail: hhs@virginia.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_3,

Gleasonian succession theory	An explanation of succession emphasizing the importance of the attributes of individual organisms as the fundamental basis.
Indicator species	Plants or animals whose presence implies the past or present conditions at a given location.
Monoclimax	The concept that one mature vegetation will eventually be produced by successional processes in a given region.
Polyclimax	The concept that multiple mature, stable-vegetation types can develop from the successional processes operating in a given region.
SIGMATIST	A school of vegetation science emphasizing the classification of vegetation.
Time-for-space substitution	The collection of vegetation data from different locations at which succession has been initiated at different times in the past to piece together the pattern of succession.

Definition of the Subject and Its Importance

"Ecological Succession" is an ordered progression of structural and compositional changes in communities toward an eventual unchanging condition, the climax community [1–3]. The term "Community" is used in two ways [4]. The "Abstract Community" refers to an abstract group of organisms that recurs on the landscape, a definition, which usually carries with it an implication of a level of integration among its parts that in extreme could be called organismal or quasi-organismal; the "Concrete Community" concept refers to the collection of organisms found at a specific place and time. These terms and their meanings are topics of significant debate among ecologists, both historically and today [5, 6]. These differences in the meaning and cause of ecological succession strongly affect the formulation of policies for ecosystems management and restoration.

Succession is important in sustainability science for several reasons: Many of the laws and regulations regarding the maintenance and restoration of terrestrial ecosystems contain assumptions that are products of different theories as to how ecological succession proceeds in a given environment. Laws might require restoration to a natural condition, which is often the vegetation that is considered the stable single vegetation type on a map that is itself based on the idea that vegetation undergoes succession a vegetation type appropriate to the climate (the climax vegetation). Further, the management of land systems often has as their basis concepts as to how succession works. For example, if one subscribes to a successional theory that abused land under an inappropriate management scheme will return to its natural state if the inappropriate management is relaxed, then one can utilize land until monitoring indicates the ecosystem is over stressed. If an alternate successional theory is applied, that abused land may not recover to its former when released from inappropriate management, then a much more cautious form of management is needed.

Introduction

Ecological succession and the dynamic change of communities is one of the pillars of modern ecology and it has considerable influence on sustainability science. The processes that drive succession and the regularly of patterns of succession have been the topic of considerable debate over the past century – due in no small part to the importance of the concept to basic and applied ecology. Succession is at its simplest level the pattern of change in ecological communities and these changes can be read on the landscape if one knows regional history or conversely can be used to interpret a landscape's history. In the former case, the historical pattern of land abandonment in the piedmont of the southeastern United States presents land elements abandoned and never reclaimed since the US Civil war in 1860, land left fallow in the First World War; farms left to ruin in the crushing economic Great Depression of the 1930s; more land left when the sons of farmers left for the Second World War; still other farm land taken out of production in agricultural programs reduce farm acreage in the 1950s. The landscape as one drives and automobile through Georgia or the Carolinas, is a kaleidoscopic montage with similar parcels of land cover types replicated on locations with similar history. In the latter case, a small patch of pine (*Pinus* sp.) trees on a slope bears witness to recovery from a past wildfire or a Beech (*Fagus*) forest of great stature indicates an area that has not burned, perhaps for centuries. The dependence of pine species on fire for regeneration or the sensitivity of the thin-barked beech to fire is written into the landscape response.

At the turn of the twentieth century, an American plant ecologist, F.E. Clements, framed many of the definitions and concepts in ecological succession and produced a theoretical system of concepts that explained the nature and function of ecological communities changing over time. Clements' definitions and approaches to understanding the dynamics of landscapes were used to develop regional systems of land management policies based on expectations, succession, and successional recovery. Clements also developed a niche-based concept of using what he called "indicator species" to evaluate the history and condition of land units [2]. For example, an overabundance of bitter or poisonous herbs might be indicators of overgrazing in a prairie systems; the occurrence of other species might indicate unusual soil chemistry and thus the presence of valuable minerals in the bedrock. The details of Clements' ideas were debated strongly by his contemporaries and other ecologists in the 1920s and 1930s found differences with parts of his successional concepts. For example, H.A. Gleason [7, 8] argued that the highly coordinated interactions in Clements concepts had at their basis the unique attributes of the individual plant species. These will be discussed below. Perhaps the most sharply pointed criticism was from the British ecologist, A.G. Tansley, who argued that the

Clements' view of succession had so clouded the meaning of the term, community, that a new term, the ecosystem [5], was needed to replace and properly formalize these ideas. It is significant in understanding the emphases in ecosystem studies to note that A.G. Tansley was a coauthor to the 1899 edition of Spencer's *Principles of Biology*, the leading English biology text of the late nineteenth century. Tansley's definition of the ecosystem and his lack of acceptance of Clements' and Phillips' holistic view that the ecological community was a coherent, organism-like entity was not a rejection of a holistic view of ecology, it was a debate among holistic ecologists [9]. These and other continuing and sometimes rancorous [6, 10] debates will be discussed in more detail below.

The development of the attributes of ecological succession has from its beginning had two poles: one Clementsian pole, which emphasizes holism, great interconnectedness of ecological systems and a successional progression toward an ultimate stable ecological system with an optimization of its attributes; a second Gleasonian pole, which sees the properties of ecological systems as deriving from the individual attributes of the organisms that comprise it and their interactions with their environment. The Gleasonian view is less dogmatic about the necessary outcomes of succession. Of course, sustainability science ultimately must take the measure of the dynamics of complex ecosystems, the natural and direction of their change under management. Debates among ecologists as to the nature of succession must be understood by scientists seeking to understand sustainability.

It is important to note that through the history of this strong Anglo-American debate on the nature of succession, there has been a continental European effort toward the goal of developing a taxonomic tradition in the issue of classifying vegetation into meaningful units, which is also an essential element of sustainable management. The so-called Zurich-Montpellier school (or as later called the SIGMATIST school) originated in the 1890s with work by C. Schröter and Charles Flahault. This effort was taken up and lead by J. Braun-Blanquet. This tradition emphasized a floristic description of vegetation and had the ultimate goal a classification of "associations" as the basic vegetation units of interest [11]. This interest in vegetation classification as a basis of mapping and prediction continues to the present. Significantly, vegetation classification is an ultimate element in organizing management plans and hence for sustainability as well.

Historical Development of the Concept of Ecological Succession

Different scientists in different locations have observed these patterns and drawn rather different conclusions about the root causes of succession and to the degree that it repeats itself over time and space. Fig. 3.1 illustrates the relationships among several ecologists who have developed what might be best described as "iconic" views of succession. The two axes in the figure are the level of holism in the theories and the emphasis on the degree of importance of feedbacks between the biological community and the environment.

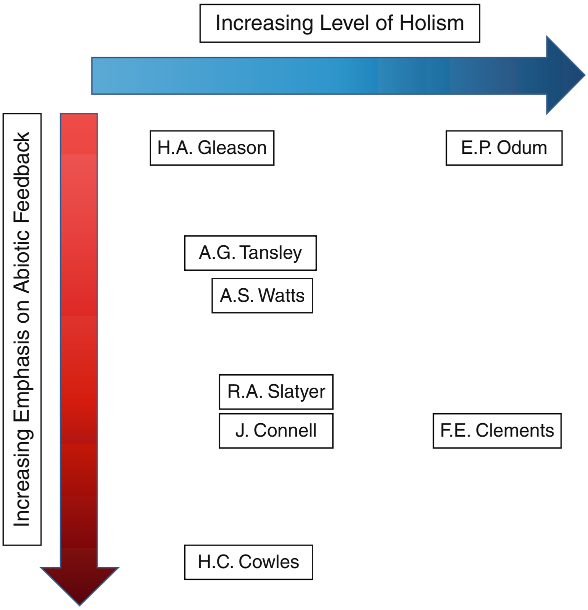


Fig. 3.1 Iconic ecologists and their emphases on the attributes of successional systems

All of these scientists are holistic in that they focused on the larger system behavior of ecological systems. The most holistic of these scientists would see the community as a highly organized, perhaps highly coevolved systems described as an abstract community in the definition section of this paper (above). They also vary on the degree to which they emphasize feedbacks between the organisms and their environment.

F.E. Clements

The iconic scientists in Fig. 3.1 represent different points of view in the formative debates on the nature of ecological succession in the formative period in the first third of the twentieth century. Clements was in some sense the dominant figure in these discussions. A.G. Tansley who represented strong opposition to many of Clements' fundamental concepts, wrote Clements obituary [12] for *The Journal of Ecology*,

> Though out of accord with many American ecologists, Clements had, throughout his career, a devoted following of younger men, and besides exercising a worldwide influence through his theory of vegetation, he directly inspired a great deal of American ecological research. This was partly due to his powerful personality. He was decidedly puritan, even ascetic (he neither drank nor smoked, and it gave him real pain to see other people doing so), and his manner was apt to be tinged with a certain arrogance. These things naturally antagonized many people; but at the same time his capacity for hard work, his intense and complete devotion to his subject, his powerful intellect and unremitting search for fresh knowledge and satisfactory formulations, could not fail to inspire the highest respect.

Clements' concept [3] saw secondary succession as being initiated by an event that removed all of the vegetation. Nowadays, this would fall under the term "disturbance" [13]. Clements referred to this process as *nudation*. Seeds, spores, etc., would then disperse to the site in the process of *migration*. These would germinate and become established in a collective set of processes that he called "eccesis." The resultant plants would then interact with one another, notably through competition. Clements called these collection of interactions, "coaction." Successful plants would interact with the microenvironment at the site and, in a process called "reaction," would modify the local site conditions as to make it unsuitable for the plant species that were there and more suitable for other species, which would replace them. This interaction is often called "facilitation" in modern ecology texts. The plant communities generated by this process are called "seres." The feedback between each sere and the environment produces an ordered sequence of seres or seral stages of succession. This reaction-replacement process would continue until an assemblage of species, which were able to occupy the site after the plant-environment reaction process, became dominant ("stabilization"). This ultimate assemblage was the "climax community."

Clements' climax community concept is incorporated almost subliminally in a number of the ideas in plant ecology and landscape management, particularly in the Western United States. Maps of "potential vegetation" are often maps of climax vegetation in the style of Clements. This is reinforced by the two-dimensional nature of paper making it cumbersome to print maps with more than one potential vegetation in a given map location. Management of parks toward a "natural" condition often has a "climax community" concept as an underlying basis. Decisions to sustainably conserve land are straight-forward if there were a climax vegetation but should be difficult for successional seres, which might be expected to change over time. While Clements' concepts are ubiquitous in land management, several ecologist have noted difficulties with the concept. More frequent criticisms of Clements' concepts include:

1. The idea that the sere/environment feedbacks produce a necessary ordering of communities toward the climax community. Clements [11] noted, "The climax formation is the adult organism of which all the initial stages are but stages of development." Some of the more ardent Clementsian ecologists [14–16] likened the sequence of seres in succession as a direct analog to the embryological development of an organism. The rejection of this concept of the community as a superorganism concept by Tansley [5] motivated the invention of the neologism, "ecosystem," as a replacement for what he felt was a corrupted term, "community," in 1935.
2. The concept of facilitation. Facilitation has been attacked from evolutionary arguments to the effect that one would not expect a species to evolve to facilitate the success of a replacing competitive species. Field observations of species replacements support the existence of facilitation such as the presence of plants with nitrogen-fixing symbiotic organisms increase the nutrient status of a site and advantaging other species. There are also numerous counterexamples of species that hold sites tenaciously over generations before finally losing to competitors [6].

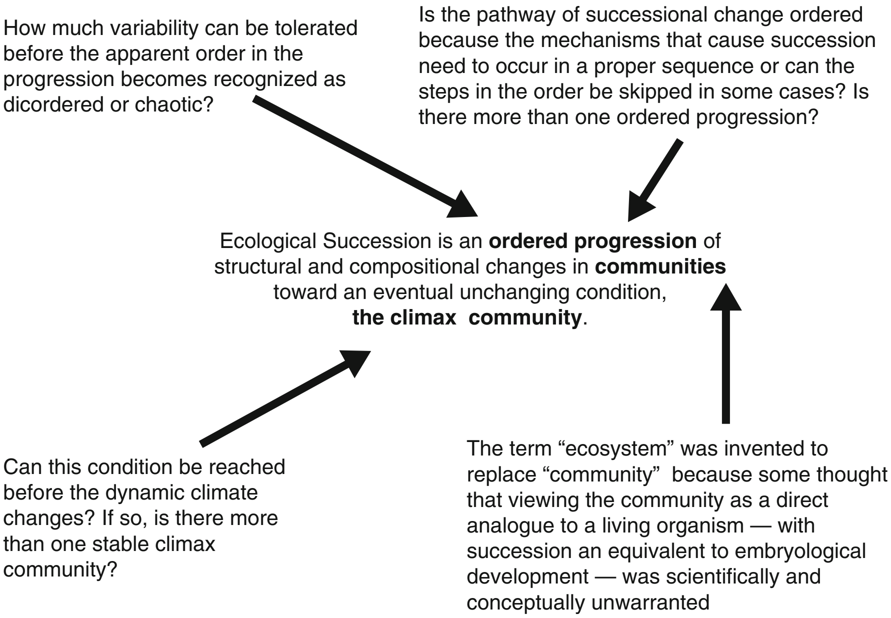
3. The progressive ecesis of different species relayed to the site for all the seral stages. In some successional sequences, all the species are present at the site from the first nudation event. In these cases, the progression of seral stages is a consequence of the differences in rate of maturation of herbaceous plants, shrubs, and trees [10].
4. The existence of a single stable climax community reached by a single sequence of seres. Several questions have arisen about the climax concept. Is there a single climax community or is there always change in the vegetation at a site? [17] Can there be multiple stable climaxes? [18–20] Can there be several pathways to a particular community? [21] Can the pathways fork? Are the seres comparable at different locations?

These criticisms continue in no small part because of the importance of Clements and his students in the formulation of a grand theory of vegetation dynamics and their ability to translate their concepts into practical applications. Subsequent recognition that past conditions could leave stable relict communities (e.g., preclimaxes such as grassland patches in temperate deciduous forest). Human or natural disturbances (such as occasional wildfires) could leave in a region a persistent vegetation that could be expected to change to a climax vegetation if the disturbance were controlled (“disclimax”). Unusual soil conditions or variations in topography could produce persistent vegetation that was different from the regional stable vegetation (“preclimaxes and post climaxes”). These elaborations of the climax concept inspired a proliferation of conditional climax vegetations [3]. These extended definitions of the Clementsian succession concept that essentially preserved the ideas of a single succession to a climax community through an ordered series of seral communities, also allowed for special conditions (disturbance regime, human use of the land, historical antecedents and soils) to create deviations from this basic model [22].

The debate over these issues and its continuation over almost a century in some cases, speaks to the complexity of the issues and the importance of understanding succession in designing concepts to manage natural systems sustainably [11, 23]. The issues raised by the scientists in Fig. 3.1 are illustrated in Fig. 3.2 as deconstructions of the standard definition of ecological succession.

H.A. Gleason and the Individualistic Concept of Succession

Henry Gleason represents a logical foil to Clements holistic idea of progression of succession toward a climax vegetation. In total, his publications were strongly focused on plant taxonomy [24] but the relatively smaller section of ecological papers that he produced had remarkable impact on his colleagues, an impact that continues to the present day. As early as 1910, Gleason [25] noted, “… it is impossible to state whether there is one definite climax vegetation in each province; it seems probable that there are several such association each characteristic of a limited proportion.” Gleason’s view of succession differed significantly from



How much variability can be tolerated before the apparent order in the progression becomes recognized as dicordered or chaotic?

Is the pathway of successional change ordered because the mechanisms that cause succession need to occur in a proper sequence or can the steps in the order be skipped in some cases? Is there more than one ordered progression?

Ecological Succession is an **ordered progression** of structural and compositional changes in **communities** toward an eventual unchanging condition, **the climax community**.

Can this condition be reached before the dynamic climate changes? If so, is there more than one stable climax community?

The term “ecosystem” was invented to replace “community” because some thought that viewing the community as a direct analogue to a living organism — with succession an equivalent to embryological development — was scientifically and conceptually unwarranted

Fig. 3.2 Some of the basic issues in the nature of ecological succession arising from a standard definition

that of Clements. He recognized as early as 1908 [26] that succession to be retrograde as well as progressive as posited by Clements. Succession was not necessarily an irreversible trend toward the climax community. Gleason also recognized that the climate could change in contrast with Clements’ theory of vegetation undergoing succession in an unchanging climate. Gleason developed what he called the “individualist concept” that succession is the result of environmental requirements of the individual species that comprise the vegetation. He also noted [27] that “. . . no two species make identical environmental demands.”

Gleason’s concept of succession was a much more fluid and much less stereotypical concept of succession than that of Clements. Succession reflected the interactions of individuals with the environment. It could change in its nature with different climatic and other environmental conditions. It could progress or regress to a more different, stable community depending on time and circumstance.

Henry Chandler Cowles

One of the founders of ecological concepts of plant communities in the United States was the geologist turned ecologist, Henry C. Cowles, at the University of Chicago [11]. Cowles [28] studied the pattern of ecological succession in a set of sand dunes, the “Indiana Dunes,” which are now protected by the State of Indiana’s “Indiana Dunes State Park.” These dunes were formed sequentially from sediment that was exposed with variation in Lake Michigan. The oldest of the dunes, which is the furthest away from the Lake Michigan beach, has a loamy brown forest soil and is covered by a Beech-Maple (*Fagus-Acer*) forest; the youngest dune at the beach front has windblown sand as a soil and is covered with a patchy beach-grassland. Cowles recognized this entire land pattern represented a chronosequence, with the spatial location of the dunes of different ages representing different stages of successional development. He later [29] produced the now famous phrase that "succession was a variable chasing a variable" – the changes in vegetation chase the changes in the climate. This insight was, and remains, a remarkable concept of vegetation function. What he meant by this was that the rate of change of vegetation succession, the variable, is sufficiently slow that one could expect the climate, the chased variable, to change by the time the succession process was completed.

Cowles was a remarkable and innovative scientist. He was honored in the 1935 issue of *Ecology*, the journal of the Ecological Society of America. The Henry Chandler Cowles issue (*Ecology*, Volume16, Number 3) is a collage of the central issues in plant ecology in its formative years: C.C. Adams, [30] an early researcher of ecological succession and the associated bird communities, along with F.E. Clements, [31] wrote on human ecology and ecology in the public service; Transeau [32] discussed the "Prairie Peninsula," a region of prairie vegetation jutting into a forests in the American Midwest, and emphasized that one must consider past climates as well as present climates to understand vegetation; Fuller [33] and Sears [34] made much the same point looking at paleoecological data; several students and colleagues of Cowles presented data on mature vegetation in different areas.

A.G. Tansley and the Ecosystem Concept

The first of the papers in the Cowles issue and different in tone from the rest, was Tansley's now classic 1935 paper on "The use and abuse of vegetational concepts and terms." Tansley [5] discussed the ideas of the American, F.E. Clements, and the amplification of these ideas by the South African, John Phillips. Tansley declared his strong disagreement with three of Phillips' papers in the *Journal of Ecology* [14–16] characterizing ecological communities as "quasi-organisms" whose successional dynamics were analogous to embryological development, "The community is born, grows, matures, reproduces, and carries out various other biotic phenomena: it behaves in a manner similar to an individual, with obvious and natural differences inherent in its wholly divergent and far more complex structure, constitution and functions. It behaves in such a way as to justify the view that not only is it similar to an organism, but that it is a kind of organism . . ." This conceptualization is in keeping with the principle of holism from Aristotle's *Metaphysica* onward and often abbreviated as "the whole is greater than the sum of its parts." Phillips attributed many of his ideas to Jan Smuts' book, *Holism and Evolution* [35] and noted later in writing a eulogy for Clements, [36] that his colleague Smuts was profoundly influenced by the writings of Clements.

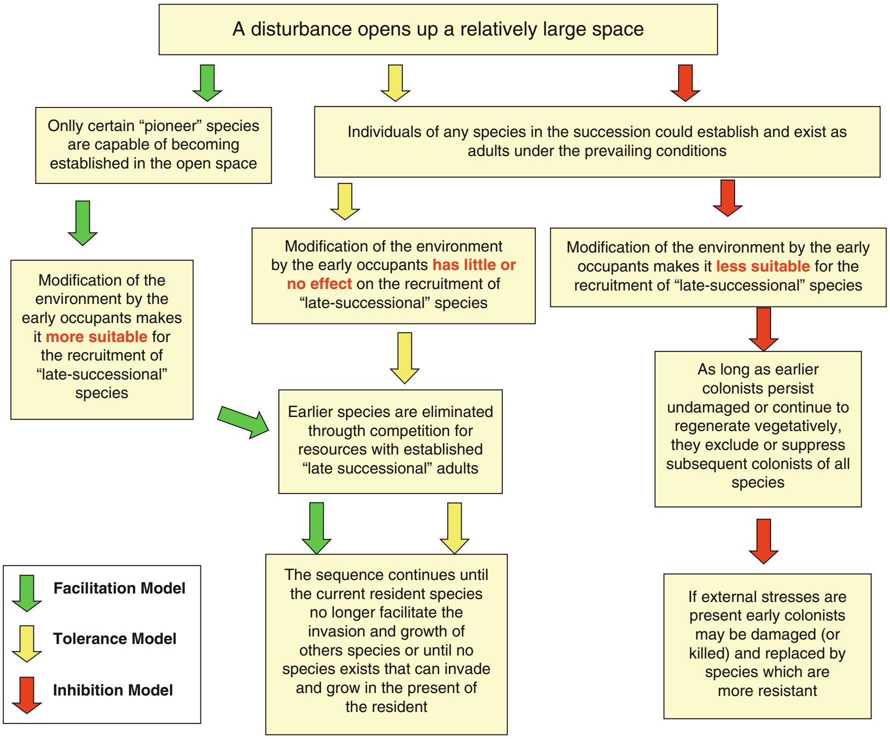


Fig. 3.3 Three different models of ecological succession from J.H. Connell and R.O. Slatyer [19] depicting succession as arising from the resident species at a location encouraging the colonization of a subsequent set of species (Facilitation Model), by the resident species being relatively neutral to the other species (Tolerance Model), or by the resident species inhibiting the colonization by other species (Inhibition Model)

Tansley created the ecosystem concept as in contrast to Phillips' and Clements' views. It is significant that in his iconoclastic "ecosystem" definition, he emphasized ecosystems "were of the most various kinds and sizes." In doing so, Tansley basically defined an ecosystem as what systems scientists would nowadays call a *system of definition*: an arbitrary system defined by the specific considerations for a particular application [37]. Tansley's ecosystem definition conforms well to more mathematical, interactive-system concepts in other sciences. His definition includes the intrinsic consideration of scale found in other sciences, particularly physically based sciences.

Tanssley also endorsed the concept of the "polyclimax", the possibility of multiple stable ecosystems following from succession and a logical contrast to the Clements idea of the "monoclimax" as the eventual product of successional processes working in a region. An example of the polyclimax is shown in Fig. 3.3 for succession after fires in the boreal forests in the vicinity of Fairbanks, Alaska. In this case, the pattern of succession following a wildfire on south-facing slopes eventually leads to a White Spruce (*Picea glauca*) forest that is quite different in its composition and character from the Black Spruce (*Picea marina*) forest seen on north-facing slopes. To some degree, Clements and his colleagues used terms such as disclimax, etc. (see above) to handle these sorts of cases in their monoclimax theory of succession.

A.S. Watt

One of Tansley's outstanding students was Alexander Stuart Watt. He developed his doctorate in 1924 on the dynamics of Beech (*Fagus* sp.) forests [38]. He noted the structure of the beech forest was at the scale of the canopy controlled by a dominant large tree best characterized as a cyclical process of the regeneration in the patch of well-lighted area that followed the death of a canopy tree, the growth of these recruits and their competition with one, and the development of a dominate canopy tree in that location. The death of this large tree would then reinitiate the cycle. By 1947, Watt had become the president of the British Ecological Society and delivered one of the most important papers in ecology as his presidential address. This presidential address and the subsequent paper represented a synthesis of what nowadays might be called the "big picture" science view of his doctoral advisor, A.G. Tansley, with Watt's own insights based of his field experience and broad ecological interests. In his address, Watt represented the plant community as a "working mechanism" of interacting plant processes such as regeneration, growth, and death on individual plants producing the broader-scale pattern of plant communities. He stated, "The ultimate parts of the community are the individual plants, but a description of it in terms of the characters of these units and their spatial relations to each other is impractical at the individual level. It is, however feasible in terms of the aggregates of individuals and species which form different kinds of patches: these patches form a mosaic and together constitute the community. Recognition of the patch is fundamental to an understanding of the structure as analyzed here" [39]. To illustrate this communality in the underlying working dynamics of vegetation, he compared a diverse array of plant communities – heathlands, grasslands, bogs, montane vegetation and forests.

As was the case with Gleason, Watt noted that one must account for the results of individual plants interacting with one another, but he also felt such an accounting was "impractical at the individual level." Twenty years later, the explosive expansion of computational power that still continues to date has made such computations more and more feasible. With the development of increasingly powerful digital computers, starting in the 1960s and continuing to the present, several different scientific disciplines (physics, astronomy, ecology) independently began to apply computers to the tasks of "book-keeping" the changes and interactions of individual entities. Forest "gap" models (which will be discussed below) were one of this class of individual-based models or IBM's that brought Watt's insights on vegetation dynamics into a form that could be projected using computer simulation techniques.

Attempts at Synthesis: Connell and Slatyer

The persistent dichotomy represented by Clementsian versus Gleasonian views of succession has consistently invited attempts to synthesis toward a unified concept. J.H. Connell and R.O. Slatyer [19] developed a synthesis that emphasized the individual attributes of plants and their feedback with their environment. Clementsian succession was a special case, which they called the "Facilitation Model" (Fig. 3.4). The replacement patterns in the Connell and Slatyer model were for sets of species, which also implies to some degree a replacement of one community with another, which also is more of a Clementsian concept. Succession was driven by transfers from one set of species as the facilitation model in which succession as arises from the resident species at a location encouraging the colonization of a subsequent set of species, as the tolerance model in which the resident set of species are relatively neutral in their interaction with other species, or as the inhibition model in which the resident species block the colonization by other species (Fig. 3.4).

Attempts at Synthesis: E.P. Odum

Eugene Pleasants Odum was a significant figure in what Robert McIntosh called the rise of ecosystems science. In 1953, he wrote the first modern ecology textbook, [40] which was the standard text for an emerging formulation of ecology as a formal academic discipline. He and his brother H.T. Odum were instrumental in the use of energy transfers in ecosystems (ecological energetics) as a common currency for comparing ecosystems with one another. As the sons of a crusading sociologist, H.W. Odum, the two Odum brothers were drawn to include "human ecosystems" in their studies, particularly the energetics of human society and its interactions with the natural ecosystems. The impact of ecological energetics on ecology was profound in that it promoted large-system intercomparisons. The International Biological Programme, the first "big picture" synoptic study of ecosystems used ecological energetics as an organizing principle for projects that compared forests, grasslands, and other biomes, worldwide. These comparisons were often made using "compartment models," diagrams in which the transfers of energy in units of such as $\text{Kcalm}^2\ \text{year}^{-1}$ were indicated by arrows connecting boxes which expressed the integral of the arrows such as Kcalm^2. Compartment models were easily transformed into ordinary differential equations for model projections

Also in the 1950s and 1960s, the United States Atomic Energy Commission (USAEC) sponsored a large program studying the movement of radioactive isotopes in natural environments including whole ecosystem experiments and direct measurements of rates of transfer of isotopes through natural environments [41]. Large, linear compartment models developed in pharmacology [42], and used in ecological energetics were developed to predict the ecological transfer of these materials. There were several advantages to such linear models. They (even for complex linear models) can be solved at equilibrium by algebraic manipulations, an

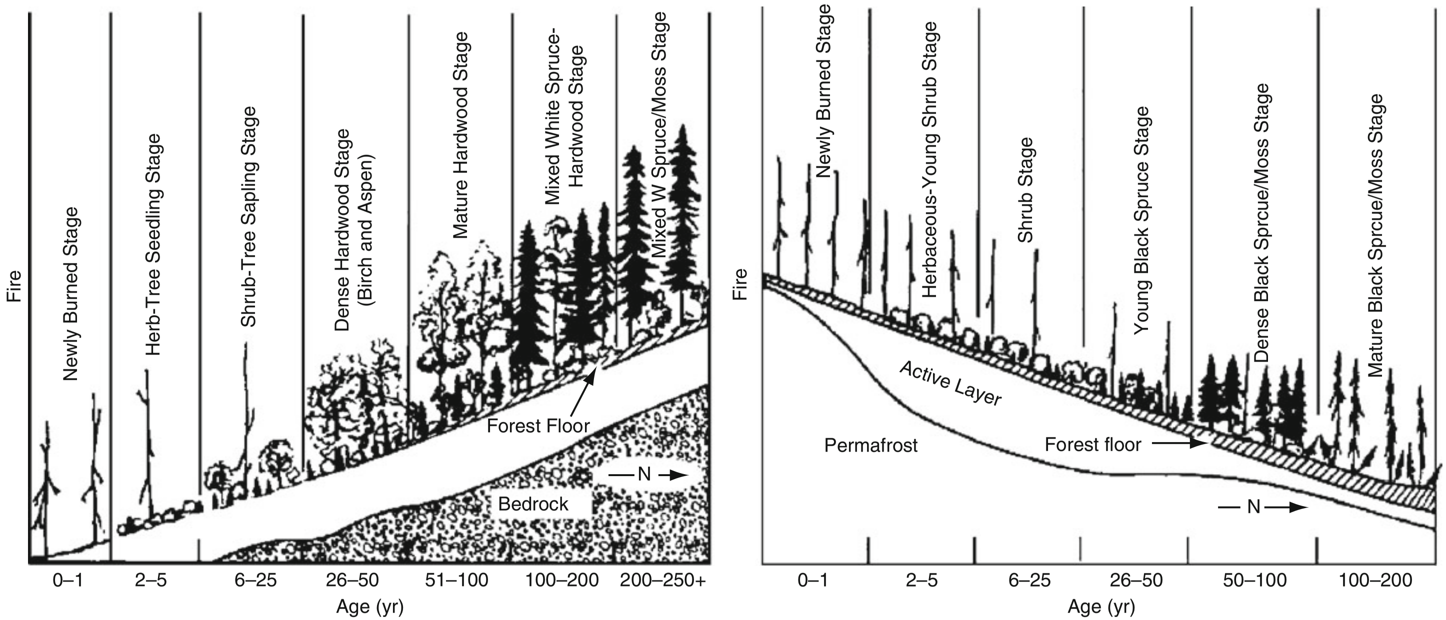


Fig. 3.4 Ecological succession following wildfires in the vicinity of Fairbanks, Alaska (USA). *Left*. Ecological succession of a south facing slope. *Right*. Ecological succession on a north- facing slope [78]

Table 3.1 A "tabular model" of changes expected over ecological succession from E.P. Odum

Ecosystem attributes	Early succession	Late succession
GPP/respiration	$>$ or <1	Approaches 1
GPP/biomass	High	Low
Biomass/energy	Low	High
Net community production (NPP)	High	Low
Food chains	Linear	Web-like
Total organic matter	Small	Large
Inorganic Nutrients	Extrabiotic	Intrabiotic
Species richness	Low	High
Species evenness	Low	High
Biochemical diversity	Low	High
Stratification and pattern	Poorly organized	Well organized
Niche specialization	Broad	Narrow
Size	Small	Large
Life cycles	Short, simple	Long, complex
Mineral cycles	Open	Closed
Nutrient exchange	Rapid	Slow
Role of detritus	Unimportant	Important
Selection on growth form	r-Selection	k-Selection
Selection on production	Quantity	Quality
Symbiosis	Undeveloped	Developed
Nutrient conservation	Poor	Good
Stability	Low	High
Entropy	High	Low
Information content	Low	High

GPP gross primary production (photosynthesis), *NPP* net primary production (GPP minus respiration)

Source: From Odum [40]

advantage given the modest computational power available in the 1950s and 1960s. When the transfer of material between compartments is a fixed percentage of the amount of material in a source compartment (called donor-controlled flow), linear models have the property that, with a constant input of material into the system, the equilibrium levels of materials are also the maximum amount. Importantly for models of radioisotope transport, radioactive decay of the isotope is a linear, donor-controlled process. Models of this formulation continue today in applications to determine the transfer of radioactive material to human populations from reactors or nuclear accidents such as the Chernobyl event.

In 1969, E.P. Odum wrote "The Strategy of Ecosystem Development," [43] a synthetic paper strongly based on ecological energetics and compartment models in ecosystems. The emphasis of the paper was strongly in the area that nowadays might be termed as sustainability science as is evidences by its first lines, "The principles of ecological succession bear importantly on the relationship between man and nature. The framework of successional theory needs to be examined as a basis for resolving man's current environmental crisis." Odum's view of succession was summarized as a tabular model of ecological succession (Table 3.1).

It was strongly Clementsian with some of the superorganism emphasis of Phillips added in. It also included a number of ideas that were popular in theoretical and population ecology at the time, such as the idea that biotic diversity in an ecosystem connotes ecosystem stability or the concept that succession was inherently driven to maximize some features (total biomass, biotic diversity, ratios of energy, or material transfers, etc.). These latter considerations can be seen as a logical continuation of Clements' concept that succession was progressive in its direction.

Odum's 1969 paper generated and still generates strong interest in the scientific community and continues to be influential these 40 years later. (At this writing in the end of year 2010, the Odum paper has been cited over 1,700 times with no diminution in citation rate.) As one might expect given its essential Clementsian theme updated with central topics in ecosystems ecology, the Odum's paper attracted several negative reviews from the ecologists of a Gleasonian persuasion, particularly given the ascendency of ideas association with Gleason's papers on succession [6]. Nevertheless, Odum's work has had a strong influence on the direction of sustainability science. Its rules on how change should progress toward appropriate natural goals provide guidelines to how ecological systems might be managed toward naturally appropriate goals. Its dogmatism makes it a target for criticism but also gives it an influential position in those wrestling with the difficult problem in a dynamical, strongly human-altered planet.

Sustainability and Vegetation Dynamics

In the historical review above, one finds succession to be a complex concept with different interpretations as to its causes and attributes over time. Ecologists such as F.E. Clements and E.P. Odum viewed succession as a process that drove change in vegetation everywhere. The generality of their application invited the inspection of local examples and the discovery of "exceptions to the rule" that could whittle away at the claimed generality. They also were drawn to applying these principles in understanding human changes on landscapes. Other ecologists such as H.A. Gleason and probably the majority of modern ecologists active in this area of research today developed a much more diffuse notion of succession. How do these ideas contribute to sustainability sciences and how can they be applied?

Clementsian Succession and Sustainability Science

The importance of succession for sustainability science can be illustrated by two examples: The first is from the historical role of succession theory in land-use policy in the United States; the second involves the incorporation of succession theory into environmental policy. Both examples favor Clementsian successional approaches in application in sustainability science. A following section will discuss the application of more Gleasonian approaches to sustainability science.

Succession and the United States Land-use Policy

The importance of succession theory in the land-use policy of the United States is can be illustrated in human connections associated with F.E. Clements. Clements' doctoral advisor at the University of Nebraska was Charles Edwin Bessey, a leading botanist of his era. Bessey came to the University of Nebraska in 1884. He was a leading college administrator; he wrote high school and college textbooks; he served as editor for the journal *Science*. He produced a remarkable set of students: the Pulitzer Prize winning novelist, Willa Cather; Frederic Clements and his wife, Edith; and Roscoe Pound, the leading jurist of his time. Pound and Clements wrote several papers together. Pound went on to the faculty at Harvard University, became Dean of the Harvard Law School, and eventually served as an advisor to President Franklin Roosevelt involved in the structuring of the "New Deal" reformulation of the structure of American government. Among the "New Deal" legislation came administrative structures and regulations for land-use policies, conservation and land protection, soil conservation, and other topics which nowadays would clearly fall in the rubric of sustainability. Pound eventually wrote an obituary/recollection for F.E. Clements [44].

The connections among these remarkable individuals from American heartlands are indicators of a deeper relationship between the ecology of Clements, his associates and students, and land-use policies in the United States. The Great Depression of the 1930s in America was acerbated by a collapse of agriculture in the "Dust Bowl" states through the middle of the nation. A complex combination of persistent land abuse, a climatic anomaly in the form of a drought, and bank collapse accompanying the collapse of family agriculture turned dryland farmers into unemployed refugees. The universities of the Midwest and the emergence of focused studies, such as the "Botanical Seminar" of Bessey at Nebraska (which had collected the Clements, Pound, and other students), had been in part a larger problem of understanding the difficulties in agriculture in the United States as farming moved westward from the forest of the East to the central grasslands [45]. Range management as an agricultural science grew from this tradition and particularly from the work of Clements and his colleagues [46, 47]. Thus, the emerging science of range management developing at the universities at the center of an ecological disaster involving the dynamics of grassland ecosystems produced information and applications on what could and should be done to restore stability. Political and policy advisors, such as Roscoe Pound, were in a position to translate this science and to put into laws concerning the sustainable use of grasslands, rangelands, and other ecosystems – in American and to a degree, worldwide.

Clarity of Clementsian Succession and the Formulation of Policy

Environmental policy like most political processes eschews complexity. One of the features of modern is the need for "elevator speeches" – summaries of complex issues that can be delivered to a listening decision-maker in the length of time one might have riding an elevator to a meeting. These and the even shorter snippets of information present to the press convey the idea that good ideas necessarily can be expressed succinctly. Certainly the Principle of Parsimony or "Ockhams Razor" in the sciences, the idea that one should select the hypothesis that makes the fewest additional assumptions when hypotheses are otherwise equal and a founding concept in modern science, lends itself in some interpreters to this point of view. Readers of F.E. Clements works are almost always impressed by the formidable richness of details in his writings, which he produced to support relatively clear general statements as to how succession as a unified process works. E.P. Odum, in some sense an inheritor and elaborator of Clements concepts, wrote using rich analogies to convey simple principles that operate in complex systems.

Clements' and Odum's writings share the ability to convey complexity with simple principles has served to make succession a tangible concept for policy makers and to illustrate the application of these principles. These "ecological generalizations as principles" have been widely used to impress upon the general public an appreciation of a unified ecological view of succession. The seemingly less-focused, special-case orientation of Gleason and subsequent ecologists seems complicated by exceptions. Range management research, which today still has a Clementsian theme and which has been successful in producing widely applied and useful policies for sustainable land management, is a tangible example that this approach can work. It is being challenged with its application in developing nations and in areas with old soils and variable climates but it remains the central paradigm [48].

Even though most modern ecologist would see Gleasonian views of succession as most appropriate, the findings do not always lend themselves to the "elevator-speech" test. One might ask, "What does an application of Gleasonian succession theory in sustainability look like in its application?" The answer is in modern forest and the very root of the origin of the word sustainability. This is the topic of the following section.

Nachhaltigkeit, Sustainability and Modern Forestry

Throughout Europe in Medieval times there was a substantial clearing of forests and extensive regional deforestation [49]. As forests, conflicts over the use products of forests intensified. This eventually produced a class conflict and subsequently laws against poaching of animals, thieving of wood, and proscriptions against public use of forests, in general. In the mid-eighteenth century, a forest management concept called “Nachhaltigkeit” (translated into English as the neologism, “sustainability”) was developed by the Germans. From about 1,800, this new forestry practice spread over Europe, particularly Northern Europe. Nachhaltigkeit involved detailed determination of how to best manage forests to produce wood and other goods in perpetuity. In German, the root Nachhaltigkeit alone means nothing more than having a lasting or sustained effect.

Nachhaltigkeit involved detailed determination of how to best manage forests to produce wood and other goods. Essential to this objective was manipulating density by spacing trees on a given site, either by planting trees or by thinning a naturally regenerated stand of trees following a timber harvest or a natural disturbance. Additionally, one needed to determine how long one should wait before harvesting a stand of trees and then planting a new stand. This spacing/length-of-rotation problem had long been solved for crop plants through experimentation and observation. To produce trees as long-lived crops, elaborate long-term data collection started on the height; size in diameter; amount of wood; size of crowns in forest stands of different densities at sites with different environmental conditions. Eventually, a forest modeling concept called the “yield table” approach developed and became the signature of modern forestry. Nowadays, some of the historical forest data-sets used in yield tables have grown to 200+ years of continual record [50].

On a given kind of site (same soils, same rainfall, etc.), trees growing in even-aged, single-species stands (such as forest plantations) tend to grow to the same height at a given age, regardless of density of trees [51, 52]. At low densities, trees have with large diameters and crowns and on an equivalent high-density location, the trees would have small diameters and crowns – but the heights of the trees would be the same in both cases. In a yield table, decades and sometimes centuries of forest stand data are arranged by the height of the trees at a given location reach at a given age, usually the typical age of tree harvest. The tree height at this standard age is called a “site index” and is used to signify the overall quality of a location for growing trees. A site index is the canopy height one would expect a forest to attain in a reference length of time. If, for example, the dominant trees in the canopy of a forest attain a height of 30 m in 50 years (the reference time) then the forest would have a site index of 30. Sites with larger site indices would reach the 30 m height earlier than 50 years; sites with lower site indices would reach 30 m later (if at all). Using yield tables, one can assess the volume of wood in managed stands of equivalent heights on different sites.

Site index is clearly defined in terms of the basic data that goes into a yield table and it can be directly determined by measuring the heights of trees on even-aged stands at the reference age. Associating site index with actual plots of land is a learned skill and an art at the same time. A capable site surveyor can judge site index by reconnaissance of land in a particular region and can make a good wage practicing this trade. Along with such arcane practices as ax-throwing, log-rolling, and tree-felling, forestry schools have regular intercollegiate competitions of judging site indices among their students as part of “Forestry Field Day” celebrations. At the edge of most universities with a college of forestry, one finds plots of trees planted at different densities for field teaching on the calibration of yield tables. The yield table concept is the quantitative basis of modern forestry.

One significant observation found in yield tables is that the rate of mortality of trees in stands undergoing natural thinning is highest in stands with the highest site indices. Faster growing trees suppress their competitors at a higher rate than slower growing trees on low site-index locations. The trade-off between rapid tree growth versus high mortality on good sites is the basis of the development of thinning strategies for managed stands and for schemes to plant forest plantations at particular spacing of trees. Despite their elevated mortality, sites with high site indices are more productive than low site-index locations.

Nachhaltigkeit was fueled by social unrest. It resulted in laws and regulation based on scientific research and experimentation with inspiration stemming from a shared political necessity to better manage the forests of Europe. Nachhaltigkeit was also the origin of the initial use in English of the word, sustainability. Its development into modern forestry practice was an ongoing process that continues today. Modern forestry is significant in sustainability science as a first real-world application. Modern forestry and its Nachhaltigkeit origins also represent a model of Gleasonian approaches to forest dynamics into sustainability science. It focuses on individual organisms and their relationships with one another and with their environment. One of the significant implications is that quantitative ecology and in particular, ecological models, likely could have a significant role in sustainability science. Individual-based vegetation dynamics models have a logical intellectual inheritance from forestry studies and are an example of tools for the application of Gleasonian successional concepts. These are the topics of the next section.

Individual-Based Models: Computer Descendants of Gleasonian Succession

Currently, several factors are simultaneously arising that challenge the application of successional theory in sustainability science. Climate conditions are changing, perhaps with a greater rate of change than seen in the past [53]. The carbon dioxide composition of the atmosphere has been altered by human activities with potential effects on plant processes particularly photosynthesis and the efficiencies of plant water-use [54]. This issue was addressed several decades ago by foresters who realized that the elaborate calibrations used to develop long-term data on forest change were subject to less accurate predictions if selected genetic strains of trees were used, if forests were fertilized or if there were significant variations in climate.

Impressed by the power of computers and interested in developing a methodology to use highly detailed computer models to simulate the spatial relations among thinning of trees, tree growth, and spatial arrangement of trees, foresters developed individual-based dynamic models of forests (mostly for commercial forests) in the mid-1960s. Their computational innovation was being independently paralleled in other fields, notably astronomy, physics, and several engineering sciences [55, 56]. The early individual-tree-based forest models were quite complex. For example, the competition among individual trees was typically simulated by crown interactions involving the 3-dimensional geometry of each individual tree crown for all the trees in a stand; [57] the growth of the tree trunks of each tree were often simulated at multiple heights; [58, 59] a tree's mortality was related to 2- or 3-dimensional crown pruning among trees [60].

For vegetation dynamics, the earliest individual-based model of forests [57] was developed for Douglas-fir (*Pseudotsuga menziesii*) applications and there were soon several similar models developed as doctoral dissertations at several Schools of Forestry [37]. These early individual-based models of forests were inordinately complex, perhaps because they were applied to tree-spacing problems in even-aged and single-species forests or forest plantations. However, a most detailed early model [61] with multiple tree species, tree interactions based on interacting three-dimensional crown geometry, mapping exact locations of trees, calculating the dispersion of seed-fall based on seed morphology, etc., exceeds the complexity of almost all of the individual-based models of today.

As noted above, Watt noted that to understand succession one must account for the results of individual plants interacting with one another, but that it is "impractical at the individual level." Twenty years later, the explosive expansion of computational power that still continues to date has made such computations more and more feasible. Early versions of these models in ecology were developed by population ecologists interested in including animal behavior in population models [62–64] and lead to a diverse array of applications for fish, insects, and birds [55]. An advantage of such models is that two implicit assumptions associated with traditional ecological modeling populations are not necessary, namely that:

1. The unique features of individuals (including their size and relative location) are sufficiently unimportant to the degree that individuals are assumed to be identical and
2. The population is "perfectly mixed" so that there are no local spatial interactions of any important magnitude [55].

As Watt noted, most ecologists are interested in variation between individuals (a basis for the theory of evolution and a frequently measured aspect of plants and animals) and appreciate spatial variation as being quite important. Assumption that this variation somehow is uniform seems particularly inappropriate for trees which are sessile and which vary greatly in size over their life span. This may be one of the reasons why tree-based forest models are among the earliest and most widely elaborated of this genre of models in ecology.

In 1972, Botkin and colleagues [65] produced an important simplification of the early forestry work (called the JABOWA model for the initials of its developers) and introduced this modeling approach to ecologists. They did not cite any of the antecedent forestry models and likely were not aware of them. The earlier forestry work remained sub rosa to most ecologists. The forestry models were mostly, but certainly not entirely, in sources not frequently read by forest ecologists. Additionally, some of the forestry models were seen as sufficiently useful that the models drifted into the realm of "industrial secrets" as their developers found employment in the timber industry.

In 1980, a paper reviewing some of this work [66] coined the term, "gap model," to describe this class of models. The gap-model designation was originally developed to emphasize that a principal simplifying assumption in these models (the assumption that the competition among individual trees on a small patch of land was homogeneous in the horizontal over a small area of land but spatially explicit in the vertical dimension) fitted well with the classic "gap dynamics" concept of A.S. Watt [39, 67]. At the time the term gap model was coined, Watt's concept had been significantly reintroduced to American ecologists by Frank Bormann and Gene Likens [68, 69]. A generation of forest ecologists have made numerous extensions of "gap models" and the term, nowadays, refers to a broad class of individual-based models of forests and other ecosystems of a natural character (mixed-age, mixed-species, natural disturbance regimes, etc.).

An Application of an Individual-Based Model on Sustainability of Russian Forests under Climate Change

Global climate model simulations indicate that the Northern Hemisphere's boreal forests and, in particular, the Siberian boreal forest zone, may not only respond to climate change but may affect the Earth's climate through feedbacks involving changes in the regional surface albedo, the degree to which the surface reflects incoming radiation. Bonan and his colleagues [70] altered surface albedo in order to simulate the clearing of the boreal forest in the National Center for Atmospheric Research (NCAR)'s Community Climate System Model version 1.0 (CCSM1). This substantially cooled the Earth not only in the boreal zone but across the Northern Hemisphere. Betts [71] used the Hadley Center Atmosphere Model (HadAM3) to simulate the climatic consequences of albedo changes from growing more trees, worldwide. He found that the surface albedo changes associated with the growth of coniferous evergreen trees in boreal regions led to significant increases in the average global temperature. These increases were large enough to overshadow the effect of the carbon storage that occurred as a result of growing evergreen forest in that region.

Field observations provide further evidence that changes in the boreal forest may impact the global climate. Larch forest, dominated by both *Larix sibirica* and *L. gmelinii*, covers extensive regions in Siberia. Shifts from larch to dark-conifer forests, dominated by trees such as spruce or fir that are tolerant of higher temperatures, have been documented [72, 73]. Because larch is a deciduous conifer, this shift in forest composition would lead to the similar albedo changes as the evergreen tree growth simulation presented by Betts. This reduction of albedo associated with a documented shift in forest type (larch to dark conifer) indicates that warming temperatures may lead to a positive feedback response: a warmer climate accelerates the natural succession from larch to dark-conifer forest; the resultant albedo change promotes additional warming.

Dynamic vegetation models, specifically forest gap models, are ideally suited to an exploration of the impacts that climate change may have on the structure and composition of boreal forests and the existence of a climate/land cover feedback in this region. The FAREAST [74] model was run at a total of 2,083 sites across the former USSR. FAREAST uses monthly climate parameters derived from historical station data to compute daily temperature and update soil water [75]. In particular, at each site, the model's climate inputs are drawn from a statistical distribution of monthly values for minimum and maximum mean temperature and precipitation which is derived from 60 years of data recorded at local weather stations [76]. The model also uses values for soil field capacity and soil carbon and nitrogen [77].

The birth, growth, and eventual death of individual trees are determined in response to local site parameters such as soil moisture and nutrient availability, which are updated annually with bio-environmental conditions, soil moisture, and available nutrients. Individual trees compete for light and nutrients with stochastic processes governing the birth and death of trees in a circular twelfth hectare plot, which approximates the size of a mature tree crown. Forty-four individual tree species are included in FAREAST simulations. These species represent the genera which dominate Northern Eurasian forests. Each species is characterized by 25 parameters which describe the species' fundamental silvics and determine which species has an advantage in terms of competition for light or nutrients, or tolerance to lack of water. At each of the 2,083 sites, 200 independent twelfth hectare plots were simulated and then the modeled biomass values were averaged for each species in each year of the model run.

The overall response of Russian boreal forests to climate change when the effects of changes in temperature and precipitation are separated show higher average precipitation leads to increased biomass (Fig. 3.5a), lower average precipitation results in decreased biomass (Fig. 3.5b), and warming causes decreases in biomass for certain regions, though in parts of Siberia, where average temperatures are extremely low, warming induces an increase in forest biomass (Fig. 3.5c). There are also genera specific patterns in the shifts in biomass that occur across Russia. Specifically, there are different patterns of change for *Larix* spp. and *Pinus* spp. in response to temperature warming. Both genera display a decrease in biomass in western and southwestern Russia and the Russian Far East. The number of sites that experience a biomass decrease for *Larix* spp. is larger than the number of sites that show a decline in *Pinus* spp. In particular, the sites that show a decline in *Larix* spp. extend further northward in both European Russia and the Russian Far East. A more detailed analysis is required to determine whether these patterns are the result of a replacement of *Larix* spp. with *Pinus* spp.

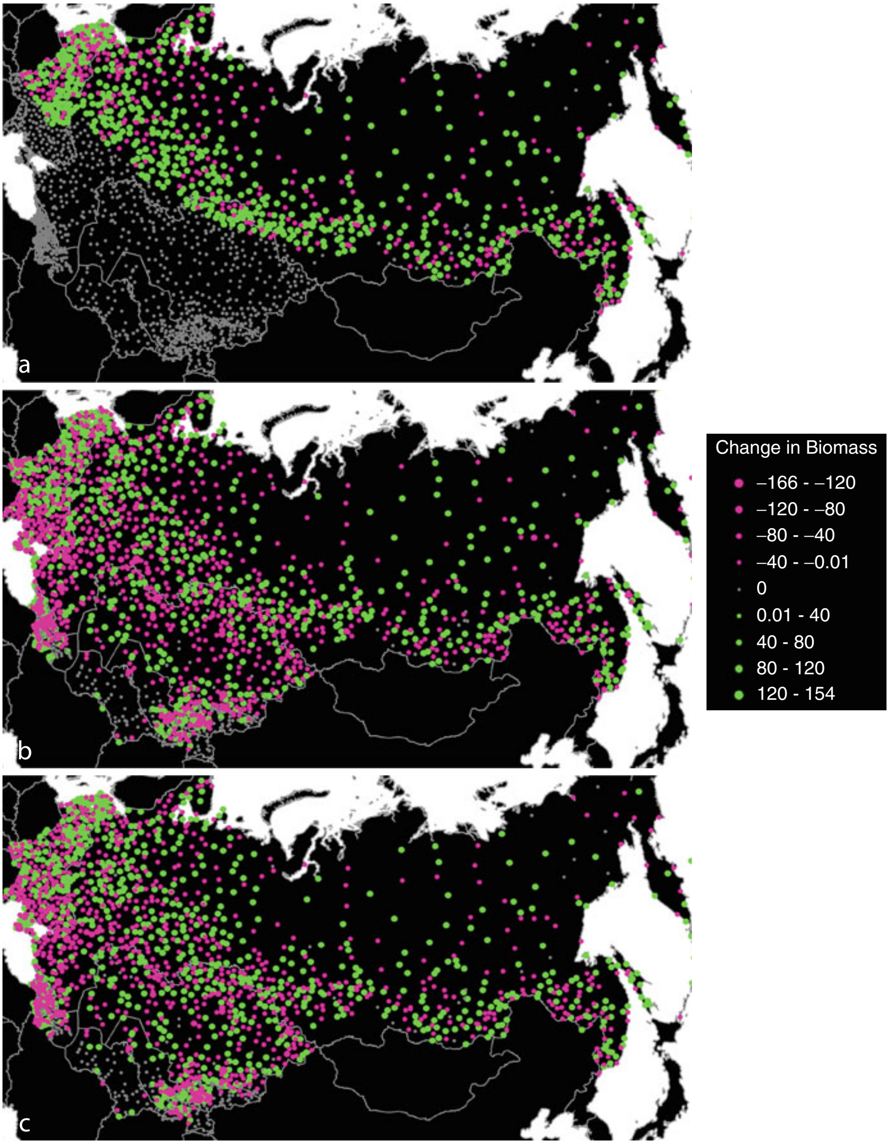


Fig. 3.5 Decreasing biomass is shown in *pink* and increasing biomass is shown in *green*. Figure shows difference in magnitude of total forest biomass (tCha^{-1}) for year 200 between a baseline scenario with no change in climate, and: (**a**) climate scenario with a +10% precipitation showing an overall pattern toward increasing biomass; (**b**) climate scenario of −10% precipitation with pattern is toward decreasing biomass; (**c**) climate scenario with a 2°C increase with decreasing biomass in western and southwestern locations, but an inconsistent response across the region

Future Directions

Ecological succession has a complex and controversy rich history. It is significant over this history that the importance of successional knowledge as a prerequisite to the wise management of landscapes has not been a point of issue. The debates in succession are not about its importance. Rather, the debates were and are over how succession works, how it should be understood, and how it can be wisely applied. This is all the more so in the case of sustainability science as a user of successional theory and succession concepts. To sustainably manage the planet, it is an absolute prerequisite to have a capability to predict the dynamic change of the world's vegetation. To the extent that the understanding of succession and vegetation dynamics is not complete, so too will be the state of the understanding of sustainable management.

Ecological models of vegetation dynamics have developed to a great degree over the past several decades. The capability to predict change in vegetation is much more today as it was 50 years ago. At the same time, the novel and large environmental changes that humankind is visiting upon the biosphere (changes in the gasses of the atmosphere, in the diversity of regions, in the local to global climate, etc.) conspire to challenge models to predict outside the range of conditions in which they were developed. Extrapolation is always a difficult endeavor for life scientists.

One would expect sustainability scientists to become more dependent on models when conditions are novel. This novelty conspires to make models potentially less reliable. Sustainability has a word origin in the word, Nachhaltigkeit. That legacy grew into modern forestry with a quantitative capability to predict the future changes of forests under active management. In some senses, however, forestry with its emphasis on single-species, even-aged forest stands has developed the simple case and provided an example of how much effort it takes to develop such the simple case. The task of understanding the complex cases of natural ecosystems in the face of novel conditions is both daunting and essential. The real challenges to sustainability science is the development of meaningful policies and protocols when the essentials, such as succession theory, are themselves changing with new discovery.

Bibliography

Primary Literature

1. Clements FE (1916) Plant succession: an analysis of the development of vegetation. Carnegie Institute, Washington, DC, Publication no 242
2. Clements FE (1928) Plant succession and indicators. Wilson, New York
3. Clements FE (1936) Nature and structure of the climax. J Ecol 24:252–284
4. McCune B, Grace JB, Urban DL (2002) Analysis of ecological communities. MjM Software Design, Gledenon, 304 pp

5. Tansley AG (1935) The use and abuse of vegetational concepts and terms. Ecology 16:284–307
6. Drury WH, Nisbet ICT (1973) Succession. J Arnold Arbor 54:331–368
7. Gleason HA (1926) The individualistic concept of the plant association. Bull Torrey Bot Club 53:1–20
8. Gleason HA (1939) The individualistic concept of the plant association. Am Midl Nat 21:92–110
9. Rich PH (1988) The origin of ecosystems by means of subjective selection. In: Pomeroy LR, Alberts JJ (eds) Concepts of ecosystem ecology. Springer, New York, pp 19–28
10. Egler FE (1954) Vegetation science concepts. I. Initial floristic composition – a factor in old-field vegetation development. Vegetatio 4:412–417
11. McIntosh RP (1985) The background of ecology: concept and theory. Cambridge University Press, Cambridge, 383 pp
12. Tansley AG (1947) Obituary notice. Frederic Edward Clements, 1874–1945. J Ecol 34:194–196
13. White PS (1979) Pattern, process and natural disturbance in vegetation. Bot Rev 45:229–299
14. Phillips J (1934) Succession, development, the climax and the complex organism: an analysis of concepts. I. J Ecol 22:554–571
15. Phillips J (1935) Succession, development, the climax and the complex organism: an analysis of concepts. II. J Ecol 23:210–246
16. Phillips J (1935) Succession, development, the climax and the complex organism: an analysis of concepts. III. J Ecol 23:488–508
17. Heinselman ML (1981) Fire and succession in the conifer forests of northern North America. In: West DC, Shugart HH, West DC (eds) Forest succession: concepts and application. Springer, New York, pp 374–405
18. Cattelino PJ, Noble IR, Slatyer RO, Kessell SR (1979) Predicting the multiple pathways of plant succession. Environ Manage 3:41–50
19. Connell JH, Slatyer RO (1977) Mechanisms of succession in natural communities and their role in community stability and organization. Am Nat 111:1119–1144
20. Glenn-Lewin DC, Peet RK, Veblin TT (eds) (1992) Plant succession: theory and prediction. Chapman and Hall, London
21. Kessell SR, Potter MW (1980) A quantitative succession model for nine Montana forest communities. Environ Manage 4:227–240
22. Weaver JE, Clements FE (1938) Plant ecology, 2nd edn. McGraw-Hill, New York, 520 pp
23. McIntosh RP (1981) Succession and ecological theory. In: West DC, Shugart HH, Botkin DB (eds) Forest succession: concepts and application. Springer, New York, pp 10–23
24. McIntosh RP (1975) H.A. Gleason – "Individualistic ecologist" 1882–1975; His contributions to theoretical ecology. Bull Torrey Bot Club 102:253–273
25. Gleason HA (1910) The vegetation of the inland sand deposits of Illinois. Bull Illinois State Lab Nat Hist 9:21–174
26. Gleason HA (1927) Further views on the succession concept. Ecology 8:299–326
27. Gleason HA (1917) The structure and development of the plant succession. Bull Torrey Bot Club 44:463–481
28. Cowles HC (1899) The ecological relations of the vegetation on the sand dunes of Lake Michigan. Bot Gaz 27:95–117, 176–202, 281–308, 361–369
29. Cowles HC (1901) The physiographic ecology of Chicago and vicinity. Bot Gaz 31:73–108
30. Adams CC (1935) The relation of general ecology to human ecology. Ecology 16:316–335
31. Clements FE (1935) Experimental ecology in the public service. Ecology 16:342–363
32. Transeau EN (1935) The prairie peninsula. Ecology 16:423–437
33. Fuller GD (1935) Postglacial vegetation of the Lake Michigan region. Ecology 16:473–487
34. Sears PB (1935) Types of North American pollen profiles. Ecology 16:488–499
35. Smuts JC (1926) Holism and evolution. Macmillan, London

36. Phillips J (1955) A tribute to Frederic E. Clements and his concepts in ecology. Ecology 35:114–115
37. Shugart HH (1998) Terrestrial ecosystems in changing environments. Cambridge University Press, Cambridge, 537 pp
38. Watt AS (1925) On the ecology of British beech woods with special reference to their regeneration. II. The development and structure of beech communities on the Sussex Downs. J Ecol 13:27–73
39. Watt AS (1947) Pattern and process in the plant community. J Ecol 35:1–22
40. Odum EP (1953) Fundamentals of ecology. W. B. Saunders, Philadelphia, 383 pp
41. Jordan CF (1986) Ecological effects of nuclear radiation. In: Orians GH (ed) Ecological knowledge and environmental problem solving: concepts and case studies. National Academy Press, Washington, DC, pp 331–344
42. Zilversmit DB, Entenmann C, Fishler MC (1943) On the calculation of turnover time and turnover rate from experiments involving the use of labeling agents. J Gen Physiol 26:325–331
43. Odum EP (1969) The strategy of ecosystem development. Science 164:262–270
44. Pound R (1954) Frederic E. Clements as I knew him. Ecology 35:112–113
45. Arthur JC (1895) Development of vegetable physiology. Science 44:164–184
46. Tobey R (1976) Theoretical science and technology in American ecology. Technol Cult 17:461–534
47. Tobey R (1981) Saving the prairies: the life cycle of the founding school of American plant ecology 1895–1955. University of California Press, Berkley
48. Washington-Allen RA, West NE, Douglas Ramsey R, Phillips DK, Shugart HH (2010) Retrospective assessment of soil stability on a landscape subject to commercial grazing. Environ Monit Assess 160:101–121
49. Williams E (2003) Deforesting the earth: from prehistory to global crisis. University of Chicago Press, Chicago, 689 pp
50. Pretzsch H (2009) Forest dynamics, yield and growth: from measurement to model. Springer, Berlin
51. Kimmins JP (1987) Forest ecology. Macmillan, Vancouver
52. Kira T (2001) Forest and environment: an approach to global environmental issues. Shin-Shisosha, Tokyo
53. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) (2007) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK/New York, 996 pp
54. Shugart HH, Woodward FI (2011) Global change and the terrestrial biosphere: achievements and challenges. Wiley-Blackwell, Oxford, 216 pp
55. Huston M, DeAngelis DL, Post WM (1988) New computer models unify ecological theory. Bioscience 38:682–691
56. DeAngelis DL, Gross LJ (eds) (1992) Individual-based models and approaches in ecology: populations, communities and ecosystems. Chapman and Hall, New York
57. Newnham RM (1964) The development of a stand model for Douglas-Fir. PhD thesis, University of British Columbia, Vancouver
58. Mitchell KJ (1969) Simulation of growth of even-aged stands of white spruce. Yale Univ Sch Forest Bull 75:1–48
59. Mitchell KJ (1975) Dynamics and simulated yield of Douglas-fir. Forest Sci Monogr 17:1–39
60. Hegyi F (1974) A simulation model for managing jack-pine stands. In: Fries J (ed) Growth models for tree and stand simulation. Royal College of Forestry, Stockholm, pp 74–90
61. Ek AR, Monserud RA (1974) FOREST: computer model for the growth and reproduction simulation for mixed species forest stands. Research report A2635, College of Agricultural and Life Sciences, University of Wisconsin, Madison
62. Holling CS (1961) Principles of insect predation. Annu Rev Entomol 6:163–182
63. Holling CS (1964) The analysis of complex population processes. Can Entomol 96:335–347

64. Rohlf FJ, Davenport D (1969) Simulation of simple models of animal behavior with a digital computer. J Theor Biol 23:400–424
65. Botkin DB, Janak JF, Wallis JR (1972) Some ecological consequences of a computer model of forest growth. J Ecol 60:849–872
66. Shugart HH, West DC (1980) Forest succession models. Bioscience 30:308–313
67. Watt AS (1925) On the ecology of British beech woods with special reference to their regeneration. II. The development and structure of beech communities on the Sussex Downs. J Ecol 13:27–73
68. Bormann FH, Likens GE (1979) Pattern and process in a forested ecosystem. Springer, New York
69. Bormann FH, Likens GE (1979) Catastrophic disturbance and the steady state in northern hardwood forests. Am Sci 67:660–669
70. Bonan GB, Pollard D, Thompson SL (1992) Effects of boreal forest vegetation on global climate. Nature 359:716–718
71. Betts RA (2000) Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. Nature 408:187–190
72. Kharuk V, Ranson K, Dvinskaya M (2007) Evidence of evergreen conifer invasion into larch dominated forests during recent decades in Central Siberia. Eurasian J Forest Res 10:163–171
73. Kharuk VI, Ranson KJ, Sergey TI, Dvinskaya ML (2009) Response of *Pinus sibirica* and *Larix sibirica* to climate change in southern Siberian alpine forest-tundra ecotone. Scand J Forest Res 24:130–139
74. Yan X, Shugart HH (2005) A forest gap model to simulate dynamics and patterns of Eastern Eurasian forests. J Biogeogr 32:1641–1658
75. Shuman JK, Shugart HH (2009) Evaluating sensitivity of Eurasian forest biomass to climate change using a dynamic vegetation model. Environ Res Lett 4:045024. doi:10.1088/1748-9326/4/4/045024
76. National Climate Data Center (NCDC) (2005) TD-9813 daily and sub-daily precipitation for the former USSR Version 1.0. NOAA National Climatic Data Center, Asheville
77. Stolbovoi V, McCallum I (eds) (2002) CD-ROM land resources of Russia. International Institute for Applied Systems Analysis and the Russian Academy of Science, Laxenburg
78. Van Cleve K, Viereck LA (1981) Forest succession in relation to nutrient cycling in the boreal forest of Alaska. In: West DC, Shugart HH, Botkin DB (eds) Forest succession: concepts and application. Springer, New York, pp 185–221

Chapter 4
Ecosystem Engineers, Keystone Species

Sanne de Visser, Elisa Thébault, and Peter C. de Ruiter

Glossary

Connectance	The proportion of possible ecological interaction links between species that are realized.
Ecosystem engineer	An organism that creates or modifies its habitat [19].
Ecosystem functioning	The way ecosystems work related to abiotic and biotic components, such as chemicals, water, soil, microbes, plants, and animals.
Keystone species	A species that has a disproportionate effect on its environment relative to its biomass (Paine 1995).
Trophic level	The position a species occupies in a food chain.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

S. de Visser
Community and Conservation Ecology group, University of Groningen, Postbox 11103,
Groningen 9700 CC, The Netherlands

E. Thébault
CNRS, UMR 7618 “Biogéochimie et écologie des milieux continentaux”,
46 rue d’Ulm, 75005 Paris, France

P.C. de Ruiter (✉)
Biometris, Wageningen UR, Radix building, Droevendaalsesteeg 1,
Wageningen 6708 PB, The Netherlands
e-mail: Peter.deRuiter@wur.nl

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_4,

Definition of the Subject

This entry focuses on two ecological phenomena. The first is "keystone species" which is defined by Paine [1] as a species (mostly of high trophic status) whose activities exert a disproportionate influence on the patterns of species occurrence, distribution, and density in a community. The second is the concept of "ecosystem engineers" defined by Jones et al. [2] as organisms that directly or indirectly modulate the availability of resources (other than themselves) to other species by causing physical state changes in biotic or abiotic materials.

Introduction: Keystone Species and Ecosystem Engineers: Analysis of Concepts

Paine's definition of keystone species was inspired from the large effects of the removal of the carnivorous starfish (*Pisaster ochraceus*) from intertidal habitat, which reduced prey species diversity due to intense competition from mussel prey [3], and represents now a classic textbook in ecology. The original keystone species concept of Paine [1, 4] thus identified a very specific mechanism: the top-down regulation of community structure and diversity by a top predator (Fig. 4.1). The concept of keystone species has been later extended to a broader definition and now includes any species whose effect on ecosystems is disproportionately large relative to its low biomass in the community as a whole [5]. Keystone species are thus species which have large effects on communities and ecosystems through many different processes such as trophic interactions, pollination, or habitat modification [6, 7]. Examples include rabbits that can increase abundance and diversity of lizards

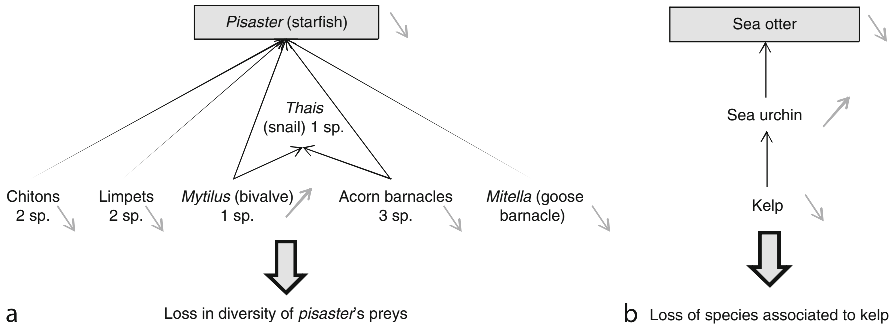


Fig. 4.1 Two examples of keystone species impacts. (**a**) Effects of the removal of *Pisaster* on prey species diversity as a consequence of mussel population explosion. (**b**) Consequences of the removal of sea otters on species diversity due to overgrazing of kelp by sea urchins. Keystone species are represented in *grey boxes*. Small *grey arrows* indicate the direction of species abundance changes following the removal of the keystone species. The large *grey arrows* indicate the global consequences of keystone species loss on the ecosystem

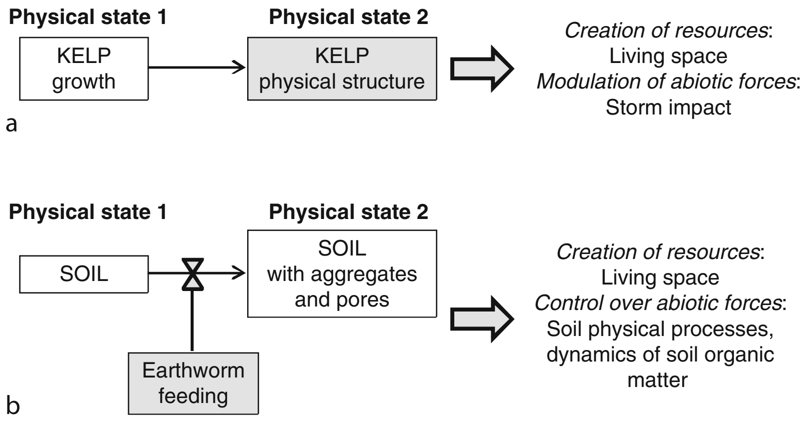


Fig. 4.2 Two examples of ecosystem engineering with kelp (**a**) and earthworms (**b**) as ecosystem engineers. Ecosystem engineering corresponds to changes in physical state (state 1–2) of biotic (i.e., kelp for **a**) or abiotic (i.e., soil for **b**) materials. Ecosystem engineers are represented in *grey boxes*. (**a**) Case of autogenic engineering, the engineer is part of the new physical state (via growth here). (**b**) Case of allogenic engineering, the new physical state is caused by the engineer (⧖, caused via feeding here), but the engineer is not part of the new physical state

[8] and sea otters whose hunting in the late nineteenth century caused a population explosion of their sea urchin prey and consequent overgrazing of kelp which led to numerous extinctions of local species [6].

The concept of ecosystem engineering was proposed two decades later than the “keystone species” concept by Jones and colleagues [2]. They defined ecosystem engineers as “organisms that directly or indirectly modulate the availability of resources (other than themselves) to other species by causing physical state changes in biotic or abiotic materials. In so doing they modify, maintain and/or create habitats” [2]. They further distinguish between two types of ecosystem engineers (Fig. 4.2): autogenic engineers that change the environment via their own physical structures, i.e., their living and dead tissues, and allogenic engineers that change the environment by transforming living or nonliving materials from one physical state to another via mechanical or other means. The idea that organisms can have important effects on abiotic processes occurring in the environment had been recognized before; indeed, Darwin devoted a whole book to the impact of earthworms on soil formation [9]. However, since the development of the concept of ecosystem engineer, engineering effects have been described for many organisms, from classic examples such as beavers, termites, or earthworms [10–12] to mollusks [13], fish [14], caterpillars [15], polychaete worms [16], grasses [17, 18], burrowing shrimp [19], ants [20], and many other species (see Table 1 of [21]).

Both the keystone species and the ecosystem engineer concepts point out to species which have important effects in ecological communities and ecosystems. Although these concepts partly overlap – an ecosystem engineer can be a keystone species – they however insist on different aspects: the keystone concept focuses on species which have disproportionate effects on community structure and ecosystem functioning ("outcome focused" sensu [22]) whereas the ecosystem engineering concept considers organisms which influence the abiotic environment with consequences on other species and related ecosystem processes ("process focused" sensu [22]). These differences between the two concepts are reflected in the literature: these concepts generally appear in distinct studies as less than 5% of the studies on these topics refer to both ecosystem engineers and keystone species (source: ISI Web of Science).

The keystone species concept has been strongly related to food web theory since its first definition [1, 23]. In particular, the identification of keystone species in food webs is an important issue. Theoretical studies have tried to pin down the characteristics of keystone species through food web models [24, 25] and several indices based on food web topology have been developed to identify keystones [26]. Models have shown that the loss of species with a large number of trophic interactions can trigger high numbers of secondary extinctions with serious consequences for species persistence; a result which highlights the potential keystone role of highly connected species in food webs [24, 27, 28].

In contrast, the concept of ecosystem engineering has been rarely related to food web studies. Recent studies acknowledge that ecosystem engineers may also play an important role in the network of trophic interactions but separating the trophic effects from the engineering effects to determine their relative importance is difficult [20, 29–31].

The importance of keystone species can also be strongly linked with ecosystem engineering. For example, the large impact of sea otters in kelp forest ecosystems results from the coupling between engineering effects and a trophic cascade [32]. In these ecosystems, kelp provides habitat for many species and dampens wave action; the keystone effect of sea otters is thus mediated through their indirect trophic effect on kelp densities which is a main ecosystem engineer.

Issue-1: How to Find Keystone Species and Ecosystem Engineers in Communities?

Keystone species and ecosystem engineers may affect ecosystem processes, such as nutrient cycling, and thereby ecosystem functioning. In the face of rapid biodiversity loss, a considerable amount of studies were dedicated to investigate a possible link between species richness and ecosystem function [33] and the threat of diversity loss on the loss of ecosystem services to man. First indications show positive relationships between species richness and ecosystem productivity, stability, and sustainability, with more species being able to fully and complementarily run ecosystem functions due to niche differentiation and facilitative interactions (reviewed by [34]). However, there is now a growing consensus that functional diversity, rather than species numbers per se, strongly determines ecosystem functioning [35]. This means that the presence of a particular species with specific traits may play a larger role in determining ecosystem function than merely the number of species [36]. The apparent diversity-ecosystem function relationship can thus be partly caused by a greater chance of an influential species with particular traits being present in more diverse communities than in species-poor communities.

If it is possible to predict and identify a priori a set of species traits that determine keystone interactions in a system, this would greatly benefit management and conservation purposes. Species' traits determine how species contribute to ecosystem processes, so the presence and distribution of such traits can be utilized to indicate aspects of ecosystem functioning [37]. To identify keystone species various methods have been used ranging from experimental removal or addition manipulations to comparative studies and natural history observations [5]. Partly because of these methodological issues, identifying keystone species has so far proved elusive [5, 38] although some progress has been made and its concept now widely investigated in the context of complex ecological networks [25, 39–41]. Some examples of specific traits are for instance trophic level, body size, connectance, or traits concerning tolerance and resilience to disturbances. Organisms that influence their environment strongly and contribute disproportionately to the functioning of ecosystems often seem to occupy higher trophic levels in food webs [5]. Top predators have been described as highly interactive keystone species [42], have been shown to play an important role in stabilizing food webs [43], and play important roles in marine ecosystems [44] and terrestrial ecosystems [45].

Also, the loss of top predators has been linked to secondary extinctions [46, 47]. This has been attributed to their ecological role as suppressors of medium-sized predators (mesopredators) (e.g., [48, 49]) and generalist herbivores [50, 51]. In terrestrial ecosystems, organisms that influence their environment strongly also often seem to be large bodied (e.g., [52]). Larger bodied organisms require a high resource and energy use per individual [53, 54] and have greater mobility, home ranges, and longevity [55, 56] and, thereby, control more resources over greater and coarser spatial scales [52, 57]. It is also proposed that well-linked and interacting species as key interactors are more important for the community [28, 58–62]. This approach characterizes the interaction structure of species placed in an ecological network. Among plants, on the other hand, some studies have shown that species within the same functional types but with different requirements and tolerances may provide insurance to the system in the form of long-term resilience against changes in environmental factors, such as global warming, grazing, drought or frost [35].

The latter example indicates that the keystone status of a species often appears to be context dependent, and may change with successional status, productivity, diversity, and other ecosystem traits [63]. It is therefore important to identify how the importance of traits that define keystone species change across a gradient of conditions, measuring environmental factors, community composition, trophic dynamics, and distribution of strong and weak links in the community (e.g., [24]). Without droughts, a specific plant species may not play an important role in maintaining community composition or ecosystem functioning. The Australian brushtail possum (*Trichosurus vulpecula*) may function as a keystone species in rata-kamahi forests by defoliating and killing canopy trees, but not in beech-dominated forests where floristic composition, but not forest structure, is typically affected [64–66]. A species in its native grounds may play no specific role in the system, whereas an invasive species may have devastating effects in the system it got introduced into, e.g., feral cats and rats on islands [67], Alewife (*Alosa pseudoharengus*) in nonnative freshwater lakes and ponds [68], and Cheatgrass (*Bromus tectorum*) in nonnative grasslands [69].

Identifying keystone species therefore is not without its problems. It is also important to notice that ecologically important species might not necessarily be the ones that are also considered important by traditional conservationists (i.e., rare species; [70]).

Issue-2: Usefulness for Management

Because of the limited resources available in comparison to conservation needs, it has been proposed to design protection of single species in the aim of indirectly protecting the regional biota. These "surrogate species" are roughly of three categories [71]: (1) flagships, charismatic species that attract public support; (2) umbrellas, species requiring such habitats that their protection might protect other species; and (3) biodiversity indicators, taxa whose presence may indicate high species richness. However the effectiveness of these policies has been questioned and [70] suggested that single-species management might be more effective when directed toward keystone species. Indeed, the importance of keystone species and ecosystem engineers in communities make these species particularly important conservation targets, since the loss of these species can affect entire communities and ecosystems [72]. However, the main difficulty for applying these concepts to conservation issues lays on both the identification of keystone species and ecosystem engineers in communities and on the context dependence of their impacts, as discussed in the previous section. Thus, although these concepts appear relevant for conservation policies, it is still a long way from providing general and practical recommendations for conservationists and managers [71].

The concepts of keystone species and of ecosystem engineers could also be useful for other management issues in natural and anthropized ecosystems, such as for ecosystem restoration or agriculture. For example, in agro-ecosystems, several well-known ecosystem engineers have been used to improve soil fertility and crop yield. In some countries, farmers make use of the soil fertilizing effect of termites by spreading termite mound soil in their field [73]. Similarly, earthworm inoculation has generally positive effects on crop yield [74].

Future Directions

The notions of ecosystem engineers and keystone species have been playing prominent roles in ecology for several decades, still many questions and uncertainties ask for further investigations. Three of them are briefly described here. The first is how keystone roles and engineering effects are related to body-size: Do larger organisms have larger effects than smaller organisms? The second regards the context dependence of keystone roles and engineering effects: As the composition and structure of ecological communities are dynamic both in terms of species composition and species abundances, what does that imply for the role species have in communities and ecosystem functioning? The third line of research might be the most relevant for our society: How can the concepts for nature conservation, biodiversity protection, and the enhancement of environmental quality be used?

Bibliography

Primary Literature

1. Paine RT (1969a) A note on trophic complexity and community stability. Am Nat 103:91–93
2. Jones CG, Lawton JH, Shachak M (1994) Organisms as ecosystem engineers. Oikos 69:373–386
3. Paine RT (1966) Food web complexity and species diversity. Am Nat 100:65–75
4. Paine RT (1969b) The *Pisaster-Tegula* interaction: prey patches, predator food preference, and intertidal community structure. Ecology 50:950–961
5. Power ME, Tilman D, Estes JA, Menge BA, Bond WJ, Mills LS, Daily G, Castilla JC, Lubchenco J, Paine RT (1996) Challenges in the quest for keystones. Bioscience 46:609–620
6. Estes JA, Palmisano JF (1974) Sea otters: their role in structuring nearshore communities. Science 185:1058–1060
7. Terborgh J (1986) Keystone plant resources in tropical forest. In: Soulé ME (ed) Conservation biology. Sinauer Associates, Sunderland, pp 330–344
8. Bravo LG, Belliure J, Rebollo S (2009) European rabbits as ecosystem engineers: warrens increase lizard density and diversity. Biodivers Conserv 18:869–885
9. Darwin CR (1881) The formation of vegetable mould through the action of worms, with observations of their habits. John Murray, London
10. Dangerfield JM, McCarthy TS, Ellery WN (1998) The mound-building termite *Macrotermes michaelseni* as an ecosystem engineer. J Trop Ecol 14:507–520
11. Eisenhauer N (2010) The action of an animal ecosystem engineer: identification of the main mechanisms of earthworm impacts on soil microarthropods. Pedobiologia 53:343–352
12. Wright JP, Jones CG, Flecker AS (2002) An ecosystem engineer, the beaver, increases species richness at the landscape scale. Oecologia 132:96–101
13. Gutiérrez JL, Jones CG, Strayer DL, Iribarne OO (2003) Mollusks as ecosystem engineers: the role of shell production in aquatic habitats. Oikos 101:71–90
14. Flecker AS, Taylor BW (2004) Tropical fishes as biological bulldozers: density effects on resource heterogeneity and species diversity. Ecology 85:2267–2278
15. Lill JT, Marquis RI (2003) Ecosystem engineering by caterpillars increases insect herbivore diversity on white oak. Ecology 84:682–690
16. Schwindt E, De Francesco C, Iribarne OO (2004) Individual and reef growth of the non-native reef-building polychaete *Ficopomatus enigmaticus* in a south-western Atlantic coastal lagoon. J Mar Biol Assoc UK 84:987–93
17. Perelman SB, Burkart SE, León RJC (2003) The role of native tussock grass (*Paspalum quadrifarium* Lam.) in structuring plant communities in the Flooding Pampa grasslands, Argentina. Biodivers Conserv 12:225–238

18. Fogel BN, Crain CM, Bertness MD (2004) Community level engineering effects of *Triglochin maritima* (seaside arrowgrass) in a salt marsh in northern New England, USA. J Ecol 92:589–597
19. Berkenbusch K, Rowden AA (2003) Ecosystem engineering – moving away from 'just-so' stories. New Zeal J Ecol 27:67–73
20. Sanders D, van Veen FJF (2011) Ecosystem engineering and predation: the multi-trophic impact of two ant species. J Anim Ecol 80:569–765
21. Wright JP, Jones CG (2004) Predicting effects of ecosystem engineers on patch-scale species richness from primary productivity. Ecology 85:2071–2081
22. Wright JP, Jones CG (2006) The concept of organisms as ecosystem engineers ten years on: progress, limitations, and challenges. BioScience 56:203–209
23. De Ruiter PC, Wolters V, Moore JC, Winemiller KO (2005) Food web ecology, playing Jenga and beyond. Science 309:68–71
24. Christianou M, Ebenman B (2005) Keystone species and vulnerable species in ecological communities: strong or weak interactors? J Theor Biol 235:95–103
25. Libralato S, Christensen V, Pauly D (2006) A method for identifying keystone species in food web models. Ecol model 195:153–171
26. Jordán F, Liu W-C, Davis AJ (2006) Topological keystone species: measures of positional importance in food webs. Oikos 112:535–546
27. Solé RV, Montoya JM (2001) Complexity and fragility in ecological networks. P Roy Soc B-Biol Sci 268:2039–2045
28. Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in foodwebs: robustness increases with connectance. Ecol Lett 5:558–567
29. Wilby A, Shachak M, Boeken B (2001) Integration of ecosystem engineering and trophic effects of herbivores. Oikos 92:436–444
30. Zhang YX, Richardson JS, Negishi JN (2004) Detritus processing, ecosystem engineering, and benthic diversity: a test of predator-omnivore interference. J Anim Ecol 73:756–766
31. De Visser SN, Freymann BP, Olff H (2011) The Serengeti food web: empirical quantification and analysis of topological changes under increasing human impact. J Anim Ecol 80:484–494
32. Jones CG, Lawton JH, Shachak M (1997) Positive and negative effects of organisms as physical ecosystem engineers. Ecology 78:1946–1957
33. Schwartz MW, Brigham CA, Hoeksema JD, Lyons KG, Mills MH, van Mantgem PJ (2000) Linking biodiversity to ecosystem function:implications for conservation ecology. Oecologia 122:297–305
34. Giller PS, O'Donovan G (2002) Biodiversity and ecosystem function: do species matter? Biol Environ 102B:129–139
35. Diaz S, Cabido M (2001) Vive la difference: plant functional diversity matters to ecosystem processes. Trends Ecol Evol 16:646–655
36. Sala OE, Lauenroth WK, McNaughton SJ, Rusch G, Zhang X (1996) Biodiversity and ecosystem functioning in grasslands. In: Mooney HA et al (eds) Functional role of biodiversity: a global perspective. Wiley, Chichester, pp 129–149
37. Bremner J (2008) Species' traits and ecological functioning in marine conservation and management. J Exp Marine Biol Ecol 366:37–47
38. Menge BA, Berlow EL, Balchette CA, Navarrete SA, Yamada SB (1994) The keystone species concept: variation in interaction strength in a rocky intertidal habitat. Ecol Monog 64:249–286
39. Brose U, Berlow EL, Martinez LD (2005) Scaling up keystone effects from simple to complex ecological networks. Ecol Lett 8:1317–1325
40. Mills LS, Soulé ME, Doak DF (1993) The keystone-species concept in ecology and conservation. Bioscience 43:219–224
41. Okey TA, Banks S, Born AF, Bustamante RH, Calvopiña M, Edgar GJ, Espinoza E, Fariña JM, Garske LE, Reck GK, Salazar S, Shepherd S, Toral-Granda V, Wallem P (2004) A trophic model of a Galapagos subtidal rocky reef for evaluating fisheries and conservation strategies. Ecol Model 172:383–401

42. Soulé ME, Estes AJ, Miller B, Honnold DL (2005) Strongly interacting species: conservation policy, management and ethics. BioScience 55:168–176
43. Rooney N, McCann K, Gellner G, Moore JC (2006) Structural asymmetry and the stability of diverse food webs. Nature 442:265–269
44. Boyd I, Wanless S, Camphuysen CJ (eds) (2006) Top predators in marine ecosystems. Cambridge University Press, Cambridge, UK
45. Johnson CN (2010) Red in tooth and claw: how top predators shape terrestrial ecosystems. J Anim Ecol 79:723–725
46. Borrvall C, Ebenman B (2006) Early onset of secondary extinction in ecological communities following the loss of top predators. Ecol Lett 9:435–442
47. Henke SE, Bryant FC (1999) Effects of coyote removal on the faunal community in western Texas. J Wildlife Manage 63:1066–1081
48. Crooks JA, Soulé ME (1999) Lag times in population explosions of invasive species: causes and implications. In: Viken A (ed) Invasive species and biodiversity management. Kluwer, Dordrecht, pp 103–126
49. Letnic M, Koch F, Gordon C, Crowther MS, Dickman CR (2009) Keystone effects of an alien top-predator stem extinctions of native mammals. Proc Roy Soc Biol Sci 276:3249–3256
50. McLaren BE, Peterson RO (1994) Wolves, moose, and tree rings on isle royale. Science 266:1555–1558
51. Ripple WJ, Beschta RL (2003) Wolf reintroduction, predation risk, and cottonwood recovery in Yellowstone National Park. Forest Ecol Manag 184:299–313
52. Owen-Smith RN (1988) Megaherbivores: the influence of very large body size on ecology. Cambridge University Press, Cambridge, UK
53. Bush GL (1993) A reaffirmation of Santa Rosalia, or why are there so many kinds of small animals? In: Lees DR, Edwards D (eds) Evolutionary patterns and processes. Academic, London, pp 229–249
54. Olff H, Ritchie ME, Prins HHT (2002) Global environmental controls of diversity in large herbivores. Nature 415:901–904
55. Haskell JP, Ritchie ME, Olff H (2002) Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. Nature 418:527–530
56. Peters RH (1983) The ecological implications of body size. Cambridge University Press, Cambridge, UK
57. Ritchie ME, Olff H (1999) Spatial scaling laws yield a synthetic theory of biodiversity. Nature 400:557–560
58. Allesina S, Bodini A, Bondavalli C (2006) Secondary extinctions in ecological networks: bottlenecks unveiled. Ecol Model 194:150–161
59. Dunne JA, Williams RJ, Martinez ND (2004) Network structure and robustness of marine food webs. Marine Ecol Prog Ser 273:291–302
60. Estrada E (2007) Characterisation of topological keystone species: local, global and "meso-scale" centralities in food webs. Ecol Complex 4:48–57
61. Jordán F, Benedek Z, Podani J (2007) Quantifying positional importance in food webs: a comparison of centrality indices. Ecol Model 205:270–275
62. Jordán F, Liu W-C, Mike A (2009) Trophic field overlap: a new approach to quantify keystone species. Ecol Model 220:2899–2907
63. De Ruiter PC, Neutel A, Moore JC (1995) Energetics, patterns of interaction strengths, and stability in real ecosystems. Science 269:1257–1260
64. Batcheler CL (1983) The possum and rata-kamahi dieback in New Zealand: a review. Pac Sci 37:415–426
65. Rose AB, Pekelharing CJ, Platt KH (1992) Magnitude of canopy dieback and implications for conservation of southern rata-kamahi (Metrosideros umbellate-Weinmannia racemosa) forests, central Westland, New Zealand. New Zeal J Ecol 16:23–32
66. Wardle JA (1984) The New Zealand beeches: ecology, utilization and management. New Zealand Forest Service, Christchurch, p 447

67. Holdaway RN (1999) A spatio-temporal model for the invasion of the New Zealand archipelago by the Pacific rat *Rattus exulans*. J Roy Soc New Zeal 29:91–105
68. Madenjian CP, Pothoven SA, Dettmers JM, Holuszko JD (2006) Changes in seasonal energy density of alewife (*Alosa pseudoharengus*) in Lake Michigan after invasion of dreissenid mussels. Can J Fish Aquat Sci 63:1–12
69. Knapp PA (1996) Cheatgrass (*Bromus tectorum* L.) dominance in the Great Basin desert. Global Environ Chang 6:37–52
70. Simberloff D (1998) Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? Biol Conserv 83:247–257
71. Boogert NJ, Paterson DM, Laland KN (2006) The implications of niche construction and ecosystem engineering for conservation biology. BioScience 56:1–9
72. Crain CM, Bertness MD (2006) Ecosystem engineering across environmental gradients: implications for conservation and management. BioScience 56:211–218
73. Black HIJ, Okwakol MJN (1997) Agricultural intensification, soil biodiversity and agroecosystem function in the tropics: the role of termites. Appl Soil Ecol 6:37–54
74. Fragoso C, Brown GG, Patron JC, Blanchart E, Lavelle P, Pashanasi B, Senapati B, Kumar T (1997) Agricultural intensification, soil biodiversity and agroecosystem function in the tropics: the role of earthworms. Appl Soil Ecol 6:17–35
75. Paine RT (1995) A conversation on refining the concept of keystone species. Conserv Biol 9(4):962–964

Chapter 5
Ecosystem Flow Analysis

Brian D. Fath

Glossary

Consumer	Heterotrophic organism that consumes other organisms for their energy requirements.
Cycling	The process by which energy or matter returns from its compartment of origin before exiting the system boundary.
Ecological goal function	Tendency observed in the orientation or directional development of ecological systems.
Flow	The transfer of energy or matter from one compartment in the system to another by active (feeding) or passive (death, egestion) means.
Network analysis	A mathematical tool to study objects as part of a connected system and to identify and quantify the direct and indirect effects in that system.
Primary producer	Photosynthesizing organism that captures external energy sources and brings it into the system as the basis for all subsequent thermodynamic activity.
Thermodynamic system	A bounded system defined by the quantities of energy and matter flowing through it.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

B.D. Fath (✉)
Biology Department, Towson University, Towson, MD, USA

Advanced Systems Analysis Program, International Institute for Applied Systems Analysis, Laxenburg, Austria
e-mail: bfath@towson.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_5,

Definition

Ecosystem services provide the basis for all human activity. Maintaining their sustained function is of critical concern to the issues of sustainability addressed here in this encyclopedia. At root, the ecosystem is a thermodynamic system receiving, collecting, transforming, and dissipating solar energy. The energy pathways are varied and complex and lead to the diversity of form and services available on the earth. Therefore, it is necessary to understand the ecosystem as a thermodynamic system and how the energy flows enter, interconnect, and disperse from the environmental system. Ecological network methodologies exist to investigate and analyze these flows. In particular, partitioning the flow into boundary input, noncycled internal flow, and cycled internal flow shows the extent to which reuse and recycling arise in ecosystems. The intricate, complex network structures are responsible for these processes all within the given thermodynamic constraints. Design of sustainable human systems could be informed by these organizational patterns, in order to use effectively the energy available. This article demonstrates the need for flow analysis, provides a brief example using a well-studied ecosystem, and discusses some of the ecosystem development tendencies which can be addressed using ecosystem flow analysis.

Introduction

Ecosystems, like all environmental systems, are open, thermodynamic systems (Fig. 5.1). They take in energy from an external energy source – almost entirely from solar energy, although geothermal or geochemical energy drives some systems. Ecosystem structure is built with the energy, and then the degraded energy is passed back to the environment. For some time period the “stored solar energy” persists in the forms perceived on the earth’s surface as biomass stores of living and nonliving organic material – such as all of us. In Frank Herbert’s novel *Dune*, he envisioned that on an arid planet an important functional role of each individual was “carrying” water in oneself. In a world dominated by thermodynamic constraints, such as ours, everyone is an energy carrier. These stores are temporary and fluid.

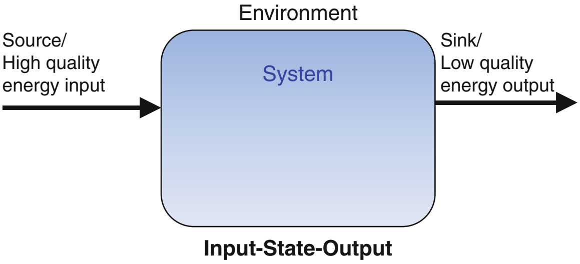


Fig. 5.1 Environmental systems are open systems, connecting to the environment through inputs of high-quality energy and discarding low-quality outputs

In this perspective, one could substitute the current object-mode paradigm with a flow-based, rheomode. In other words, *all stocks are flows.* The paper that one writes on now is only an ephemeral stage of the energy which started in the solar reactions, traveled to earth, was captured by a photosynthesizing organism, converted into storage in the xylem of that organism, and harvested and transformed into the useable product that one now holds. But that is not the end. Over time, the paper will slowly degrade or decompose, or perhaps the energy release will be sudden through combustion. In any case, the objects held are transitory states in a long-term dynamic from (energy) source to (energy) sink. Changing this view would help in appreciating better the difference between capital and income, because, for example, harvesting natural capital stock (i.e., a forest) into a flow (deforestation) is not equivalent yet treated as substitutable in the current accounts. The rheomode approach could help focus clearly on the difference and thus sustainability of stocks and flows.

Concerned about sustainability over a human time horizon, one must be aware of the constraints imposed by these sources and sinks. From the input side, clearly, it is necessary that one does not extract resources at a rate faster than they can regenerate. And concerning the output, the waste emissions should not exceed to the assimilative capacity of the local environment [1]. These are the most basic constraints imposed by open-system, thermodynamics. Humans have transformed the earth's surface to maximize the capture of photosynthetic energy – think of the millennia over which the Chinese, Romans, Babylonians, etc., have manipulated and manicured the landscape for agricultural production. Still, these societies rose and fell within the solar energy domain. These societies collapsed if they overconsumed the base resources or if they polluted their local environments [2]. In addition to these persistent input–output constraints, there is a third sustainability consideration currently observed in the anthropocene, in that it is not only the input–output relations, but also the structure which is created. Modern infrastructure demands the continual input of high-quality (low entropy) energy of a form not naturally delivered by ecosystem services. Furthermore, the created structure locks us into the necessity of immense energy flows for maintenance. Perhaps an apt analogy can be offered through the Greek myth of Erysichthon, who was King of Thessaly. He angered the gods by cutting down a sacred tree, and as punishment was insatiably hungry. Importantly, the more he ate, the hungrier he became. Our infrastructure, like Erysichthon, does not sit idle but continually demands upkeep such that the more structure, the more resources are needed to support this structure. It is not just about the present flows, but also the life cycle debt commitment as a result of the structure. Today that energy debt is paid almost entirely in the usage of fossil fuels – a nonrenewable resource. The scale of human activity seen today is because the application of fossil fuels to substitute solar fuels has released humans from one of the long-standing constraints on growth. And as a result, humans have exploded across the landscape. This growth was easy when there was sufficient energy to add to the system. In fact, the first growth form is *boundary growth*, taking energy into the system, and storing it as biomass. As long as there is more energy available, the system growth can occur unbound. The second stage of growth, *network and information growth*, is squeezing more utility out of the available source by coupling processes and improving efficiencies [3]. This can occur in parallel with boundary growth, but becomes a necessity when boundary growth is limited. In human systems, those immediate constraints are looming at least given the current fuel-mix options.

Natural ecosystems, dependent on the solar energy flows developed extremely complex and beautiful structures within these thermodynamic constraints. And it is a useful guide to learn from these systems as a more eco-friendly design is incorporated. Below, some of the ecosystem flow analysis methodologies are explored and applied.

Investigation of Ecosystem Flows Using Network Analysis

"There are no trash cans in nature." This is a useful phrase reminding that waste from one entity is food/input for another. Energy, of course, has a higher dissipative factor in the reuse than material cycles such as nitrogen, phosphorus, and calcium, but still there is a complex network of pathways designed to utilize the energy available in natural ecosystems. In 1973, Hannon [4] introduced Leontief's input–output methodology into ecology, applying it to the energy flow structure of an ecosystem. The ecosystem is represented by n compartments, and the energy flowing into compartments, within compartments and exiting compartments.

A network flow model is essentially an ecological food web (energy–matter flow of who eats whom), which also includes energy input, and nonfeeding pathways such as dissipative export out of the system and pathways to detritus. The first step is to identify the system of interest and place a boundary (real or conceptual) around it. Energy–matter transfers within the system boundary comprise the network; transfers crossing the boundary are either input or output to the network, and all transactions starting and ending outside the boundary without crossing it are external to the system and are not considered. The energy inflows and intra-system flows can be considered the *production energy flow* and the flows with no consumers such as metabolic energy and exported biomass are the *respiration energy flow* [5].

The data required for ecological network analysis are as follows: for each compartment in the network, the biomass and physiological parameters, such as consumption (C), production (P), respiration (R), and egestion (E), must be quantified [6]. Furthermore, the diet of each compartment must be apportioned amongst the inputs from other compartments (consumption) in the network. This apportionment of "who eats whom and by how much" can be depicted in a dietary *flow matrix*, **F**, where energy flows from column elements j to row elements i. For all compartments, inputs should balance outputs (C = P + R + E) in accordance with the conservation of matter and the laws of thermodynamics.

The sum of the flow matrix elements, f_{ij}, gives the total inflow to compartmental i such that:

$$T_{i,in} = \sum_{j=1}^{n} f_{ij} + z_i$$

where z_i is the boundary flow into i. The outflow from i can be expressed as:

$$T_{i,\,out} = \sum_{j=1}^{n} f_{ji} + y_i$$

where y_i is the boundary outflow from i. At steady state, a necessary condition for the network flow analysis, $T_{i,in} = T_{i,out}$ and one compartmental throughflow vector can be written as $\mathbf{T} = (T_i)$. The total system throughflow (TST) is given by the sum of the compartmental throughflows:

$$TST = \sum_{j=1}^{n} T_i$$

The motivation for flow partitioning begins with nondimensional flow intensities (i.e., throughflow-specific flows) which result when flows are divided by throughflows of originating compartments: $g_{ij} = f_{ij}/T_j$. The elements of matrix $\mathbf{G} = (g_{ij})$ give the dimensionless transfer efficiencies corresponding to each direct flow, f_{ij}. Powers $\mathbf{G}^{\mathrm{m}}$ of this matrix give the indirect flow intensities associated with paths of lengths m = 2, 3, …. Due to dissipation, flow along these indirect paths approaches zero as m → ∞ so that the power series $\sum_{m=0}^{\infty} \mathbf{G}^{\mathrm{m}}$ representing the sum of the initial, direct, and indirect flows converges to an integral flow intensity matrix, N:

$$\begin{aligned} \underbrace{\mathbf{N}}_{integral} &= \underbrace{\mathbf{I}}_{initial} + \underbrace{\mathbf{G}}_{direct} + \underbrace{\mathbf{G}^2 + \mathbf{G}^3 + \ldots + \mathbf{G}^{\mathrm{m}} + \ldots}_{indirect} \\ &= (\mathbf{I} - \mathbf{G})^{-1} \end{aligned}$$

N maps the steady-state input vector **z** into the steady-state system throughflow vector:

$$T = Nz = (\mathbf{I} + \mathbf{G} + \mathbf{G}^2 + \mathbf{G}^3 + \ldots + \mathbf{G}^{\mathrm{m}} + \ldots)z$$

Term by term, flow intensities $\mathbf{G}^{\mathrm{m}}$ of different orders m are propagated over paths of different lengths m. The first term, **I**, brings the input vector **z** across the system boundary as input z_j to each initiating compartment, j. The second term, **G**, produces the first-order direct transfers from each j to each i in the system. The remaining terms where m > 1 define m$^{\mathrm{th}}$ order indirect flows associated with length m paths. As stated before, these go to zero in the limit as m → ∞, which is necessary for series convergence. This demonstrates that each "direct" flow f_{ij} at

Table 5.1 Network representation of flow partitioning into five modes for any (i, j) pair in a system

	Pair-wise interaction	System-wide contribution
Mode 0 (boundary input)	$f_{j0}^{(0)} = z_j$	$f^{(0)} = \sum f_{j0}^{(0)}$
Mode 1 (first passage)	$f_{ij}^{(1)} = \left(\frac{n_{ij}}{n_{ii}} - \delta_{ij}\right) z_j$	$f^{(1)} = \sum\sum f_{ij}^{(1)}$
Mode 2 (cyclic)	$f_{ij}^{(2)} = \frac{n_{ij}}{n_{ii}}(n_{ii} - 1) z_j$	$f^{(2)} = \sum\sum f_{ij}^{(2)}$
Mode 3 (dissipation)	$f_{ij}^{(3)} = \left(\frac{n_{ij}}{n_{ii}} - \delta_{ij}\right) z_j$	$f^{(3)} = \sum\sum f_{ij}^{(3)}$
Mode 4 (boundary output)	$f_{0j}^{(4)} = y_j$	$f^{(4)} = \sum f_{0j}^{(4)}$

steady state is actually composed of flow elements of all orders, m = 1, 2, …. In fact, a major result of this flow analysis is that indirect flows can dominate direct flows: $\sum_{m=2}^{\infty} \mathbf{G}^{\mathrm{m}} > \mathbf{G}$. In the above developments **F**, **T**, **z**, and **y** represent matter or energy fluxes, and **G** and **N** are dimensionless intensive flows.

Finn [7] developed a cycling index using this basic approach and, Higashi et al. [8] described a three-mode partition of the flows, expanded by Fath et al. [9] into five modes (Table 5.1). Mode 0 is the boundary input into the system. Mode 1 accounts for all flow in which substance moves from node j to a terminal node i for the first time only without cycling. Mode 2 is flow cycled at terminal nodes i of each (i, j) pair. Mode 3 is component-wise dissipative flow in the sense that it exits from node i never to return again to i. Mode 4 is the boundary output from i constituting systemically dissipative flows exiting the system (Fig. 5.2). δ_{ij} is the Kronecker delta defined by $\delta_{ij} = 1$ for i = j and $\delta_{ij} = 0$ for i ≠ j.

Note, the symmetry in that quantitatively Mode 0 = Mode 4, and Mode 1 = Mode 3. This is due to the conservation of mass/energy and at steady state what comes in must go out. Mode 2 represents the cycled flow which has additional impact on the system by staying in the system longer, increasing the residence time, and returning to its source of emanation. Therefore, total system throughflow can be written as:

$$TST = f^{(0)} + f^{(1)} + f^{(2)} = f^{(2)} + f^{(3)} + f^{(4)}$$

And, on a nodal basis, throughflow is:

$$\begin{aligned} T_{ij} &= f_{ij}^{(0)} + f_{ij}^{(1)} + f_{ij}^{(2)} = z_j \delta_{ij} + \left(\frac{n_{ij}}{n_{ii}} - \delta_{ij}\right) z_j + \frac{n_{ij}}{n_{ii}}(n_{ii} - 1) z_j \\ &= \left(\delta_{ij} + \left(\frac{n_{ij}}{n_{ii}} - \delta_{ij}\right) + n_{ij} - \frac{n_{ij}}{n_{ii}}\right) z_j = n_{ij} z_j \end{aligned}$$

The mode partition designation clearly shows the contribution of flow within the entire system of interactions.

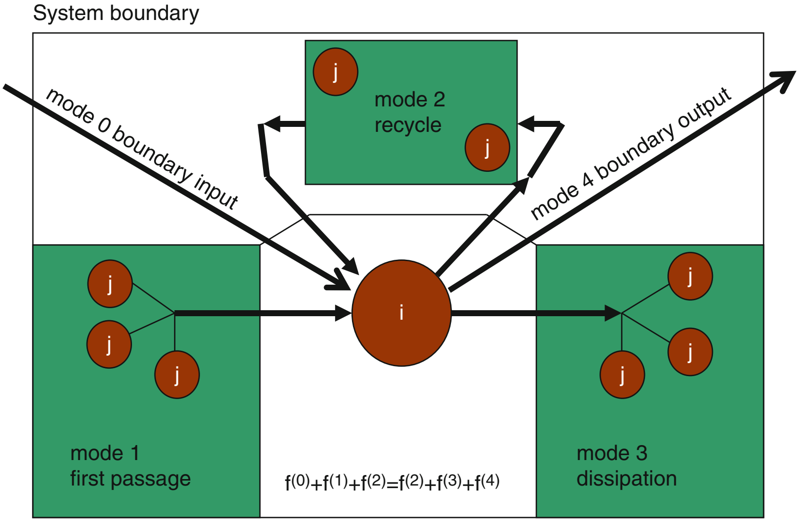


Fig. 5.2 Schematic of flow partitioning for a central node i, in relation to other network compartments. Flow reaches the node directly across the boundary, $f^{(0)}$, by passing through other compartments before reaching i, $f^{(1)}$, and leaving i to cycle back again, $f^{(2)}$. Outflows symmetrically mirror these inputs

Example: Cone Spring Ecosystem Model

A classic example is the Cone Spring ecosystem model developed by Tilly [10]. In this model, there are five compartments representing: (1) plants, (2) bacteria, (3) detritivores, (4) carnivores, and (5) detritus (Fig. 5.3). There are 2 external inputs (to plants and detritus), 8 internal flows, and each compartment has boundary outflow representing metabolic or egestion losses. The internal flows from columns j to rows i are given by:

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 8881 & 0 & 1600 & 200 & 167 \\ 0 & 5205 & 0 & 0 & 0 \\ 0 & 2309 & 75 & 0 & 0 \\ 0 & 0 & 0 & 370 & 0 \end{bmatrix}$$

Compartmental throughflows are: T = [11 184, 11 484, 5 204, 2 384, 370] and TST = 30 627.

The nondimensional flow fractions are given by:

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.794 & 0 & 0.308 & 0.0840 & 0.451 \\ 0 & 0.453 & 0 & 0 & 0 \\ 0 & 0.201 & 0.014 & 0 & 0 \\ 0 & 0 & 0 & 0.155 & 0 \end{bmatrix}$$

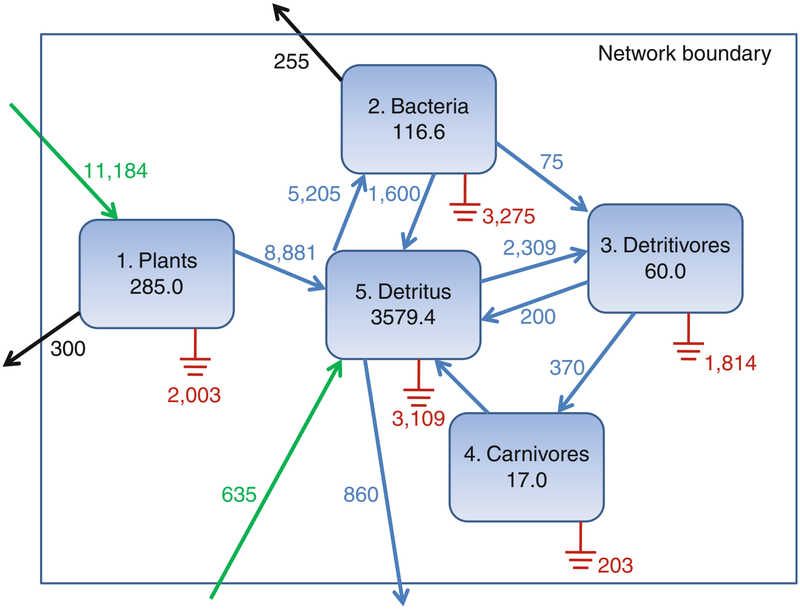


Fig. 5.3 Network diagram of Cone Spring ecosystem energy flows [10]. All flows are in kcal/m^2/year. Biomasses are in kcal/m^2. *Green arrows* are exogenous boundary inflows. *Black arrows* are exports of useable energy. *Red ground* symbols represent metabolic energy loss

Table 5.2 Results of the flow analysis partitioning for Cone Spring ecosystem. Values represent the fraction of total system throughflow for each mode

Mode 0	0.386
Mode 1	0.522
Mode 2	0.092
Mode 3	0.522
Mode 4	0.386

And the integral flow matrix is:

$$\mathbf{N} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.958 & 1.207 & 0.374 & 0.186 & 0.545 \\ 0.434 & 0.547 & 1.169 & 0.084 & 0.247 \\ 0.199 & 0.251 & 0.092 & 1.039 & 0.113 \\ 0.031 & 0.039 & 0.014 & 0.161 & 1.018 \end{bmatrix}$$

Looking at Table 5.2, it is seen that over 38% of the flow comes directly to a node from the first instant across a system boundary and 52% of the flow originates from one compartment and enters another compartment for the first time without cycling. Slightly over 9% of the total energy flow is material that has

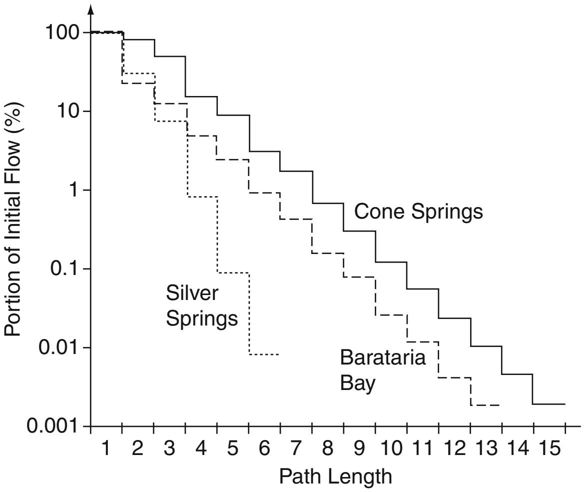


Fig. 5.4 Portion of original energy that remains as path length increases (From Braner [11], reprinted with permission). There is still noticeable energy in the system after 10 steps

cycled by exiting and reentering the same compartmental node. In other words, about 3,000 kcal/m^2/y of the total system throughflow is comprised of energy due to cyclic pathways which retain the energy in the system. This additional boost is important to the overall function of the Cone Spring ecosystem. A noticeable contribution of cycled flow is a common phenomenon in all ecosystems. Another way to demonstrate this importance of cycling, and the fundamental shift it has on how an ecosystem should be viewed, was given by Braner [11]. While investigating the same five-compartment Cone Spring model above, he showed that cyclic pathways identified by flow analysis reveal that the original boundary flow persists in the system much longer than obviously apparent. For contrast, in a five-compartment food chain model – a type often used, incorrectly, to represent an ecosystem – the longest path could only be four steps in length from X1 → $X2$ → X3 → X4 → X5. Real ecosystems have more complex structures with cycles. After those four steps, the original flow from compartment 1 would exit the system at compartment 5. According to Braner, more than 10% of the flow remains in the Cone Spring ecosystem after four steps. In fact, approximately 1% of the original flow remains after 9 steps and 0.001% is left after 15 steps (Fig. 5.4). A similar result is shown for two other ecosystems in the same figure. Therefore, the cycles, evident from flow analysis, play a very important role in the system having enough resource to function and provide ecosystem services.

Ecosystem Goal Functions

Flow analysis has another useful feature related to understanding ecosystem dynamics. Odum [12] proposed 24 different attributes which describe the ecosystem development, for aspects such as community energetics, nutrient dynamics, and overall homeostasis. The attributes dealing with energetics, which change during the ecosystem development, have also been formulated as ecological goal functions – which describe observable macroscopic patterns over time. They are not strict goal functions in the sense of mathematical optimization models (neither is economic utility although it is used as such), but indicate the tendency for ecosystems to follow during development, for example, during succession from early r-selected species and r-selecting environments to late K-selected species and K-selecting environments. Some of the more common goal functions employed include: maximum power [13], maximum dissipation [14], maximum cycling [15], maximum residence time [16], minimum specific dissipation [17, 18], maximum energy [19], and maximum ascendency [20] (see [9], for a detailed description of these). The idea is that the ecological network self-organizes itself in a way that leads to directional change in the property of these values. For example, maximum power, interpreted to mean the maximum throughflow in the network is given by: $TST = f^{(0)} + f^{(1)} + f^{(2)}$. Therefore, TST increases when there is more boundary flow (mode 0), more first passage flow (mode 1), or more cycled flow (mode 2). The mechanisms for this to increase practically relate to the system's ability to capture more boundary flow by increasing the uptake. Both first passage flow and cycled flow also depend on the second stage of growth exemplified by the structure of the network and the efficiency of flows along each connection. Similar rationale can be made for the other goal functions listed above, and in fact it has been shown that the goal functions are complementary and mutually reinforcing in that the realization of one generally promotes the others. Together they provide a holistic view of ecosystem development through different thermodynamic perspectives. Again, the value of this ecosystem knowledge is obvious for application to design and to manage human systems sustainably. If ecosystem services are required, then the inherent dynamics of the systems used should be better understood. Human activities in line with these directions will be supported by natural processes, those that do not will experience additional resistance and therefore additional cost and difficulty. Humans are better off working with nature than against it if possible.

Conclusions and Future Directions

Ecosystem flow analysis clearly shows that the distribution of energy flow in a network is not simple. Some significant fraction of the energy remains in the system and cycles before exiting the system. This insight was evident in R. Lindeman's [21] seminal work on Cedar Bog Lake in which he referred to his eight-compartment ecosystem as a "food-cycle." Unfortunately, he did not have the quantitative tools at his disposal, like flow analysis, and to simplify the calculations, proceeded to analyze the system according to two distinct "food-chains," although in reality they are linked and contain cycles. Further work in this area also neglected the presence and significance of food cycles until research in the mid-1970s (such as [22–25], and others) when network analysis techniques developed sufficiently to provide a holistic investigation of the ecosystem function. As stated above, this changes the way one must look at ecosystems, as processors and stores of energy flow. The energy does not pass quickly through but can remain and impact the system indirectly. The good news is that this flow which remains in the system is able to positively drive ecosystem processes and contribute to the overall sustainability of the system.

However, the lesson to take is that in the design of human systems, industrial processes are built sequentially, which have raw material → processing → product → disposal. There is little room for cycling and reuse. Remember, there are no trash cans in nature. Everything has a use and reuse. Efforts are now seen in industrial ecology promoting closed loop engineering and cradle-to-cradle considerations, but there is a long way to go, as evidenced by the massive amounts of raw material input and solid waste generated on a daily basis by human activity. Also, the flow analysis must include all parts and processes of the holistic integrated socio-ecological system. Future work is needed to continue to understand energy cycles in natural systems and furthermore, how to implement lessons from these into the design of socio-ecological systems. Ecosystem flow analysis clearly shows the input–output orientation flow resources have at their disposal for maintaining functional activity and can aid in sustainability science.

Bibliography

Primary Literature

1. Daly HE, Townsend KN (1993) Valuing the earth: economics, ecology, ethics. MIT Press, Cambridge, MA, p 399
2. Diamond J (2005) Collapse: how societies choose to fail or succeed. Penguin Press, New York, p 575
3. Fath BD, Jørgensen SE, Patten BC, Straškraba M (2004) Ecosystem growth and development. Biosystems 77:213–228
4. Hannon B (1973) The structure of ecosystems. J Theor Biol 41:535–546
5. Levine SH (1977) Exploitation interactions and structure of ecosystems. J Theor Biol 69:345–355
6. Fath BD, Scharler U, Ulanowicz RE, Hannon B (2007) Ecological network analysis: network construction. Ecol Model 208:49–55
7. Finn JT (1976) Measures of ecosystem structure and function derived from analysis of flows. J Theor Biol 56:363–380
8. Higashi M, Patten BC, Burns TP (1993) Network trophic dynamics: the modes of energy utilization in ecosystems. Ecol Model 66:1–42
9. Fath BD, Patten BC, Choi JS (2001) Complementarity of ecological goal functions. J Theor Biol 208:493–506
10. Tilly LJ (1968) The structure and dynamics of Cone Spring. Ecol Monogr 38:169–197
11. Braner M (1985) A simple random path method for the analysis of flow networks. Ecol Model 28:165–180
12. Odum EP (1969) The strategy of ecosystem development. Science 164:262–270
13. Lotka AJ (1922) Contribution to the energetics of evolution. Proc Natl Acad Sci USA 8:147–151

14. Schneider ED, Kay JJ (1994) Life as a manifestation of the second law of thermodynamics. Math Comput Model 19:25–48
15. Morowitz HJ (1968) Energy flow in biology; biological organization as a problem in thermal physics. Academic, New York
16. Cheslak EF, Lamarra VA (1981) The residence time of energy as a measure of ecological organization. In: Mitsch WJ, Bossermann RW, Klopatek JM (eds) Energy and ecological modelling. Elsevier, Amsterdam, pp 591–600
17. Onsager L (1931) Reciprocal relations in irreversible processes. Phys Rev 37:405–426
18. Prigogine I, Wiame JM (1946) Biologie et Thermodynamique des Phenomenes Irreversible. Experientia II:451–453
19. Odum HT (1988) Self-organization, transformity, and information. Science 242:1132–1139
20. Ulanowicz RE (1997) Ecology, the ascendent perspective. Columbia University Press, New York
21. Lindeman RL (1942) The trophic dynamic aspect of ecology. Ecology 23:399–418
22. Pomeroy LR (1974) The ocean's food web, a changing paradigm. Bioscience 24:499–504
23. Finn JT (1976) Measures of ecosystem structure and function derived from analysis of flows. J Theor Biol 56:363–380
24. Patten BC (1978) Systems approach to the concept of environment. Ohio J Sci 78:206–222
25. Patten BC (1985) Energy cycling in the ecosystem. Ecol Model 28:1–71

Books and Reviews

Burns TP (1989) Lindeman's contradiction and the trophic structure of ecosystems. Ecology 70:1355–1362

Finn JT (1980) Flow analysis of models of the Hubbard Brook ecosystem. Ecology 61:562–571

Latham LG (2006) Network flow analysis algorithms. Ecol Model 192:586–600

Ulanowicz RE (1983) Identifying the structure of cycling in ecosystems. Math Biosci 65:219–237

Whipple SJ (1999) Analysis of ecosystem structure and function: extended path and flow analysis of a steady-state oyster reef model. Ecol Model 114:251–274

Chapter 6
Ecosystem Services

Heather Tallis, Anne Guerry, and Gretchen C. Daily

Glossary

Ecosystem services	The wide array of benefits that ecosystems, and their biodiversity, confer on humanity.
Marginal value	The economic value of the next incremental unit of something. In this context, marginal values are those associated with managing the next small unit of an ecosystem in a particular way (e.g., preserving, rather than clearing, the next unit of forest). They can also be the partial contribution of natural capital to a final good that is produced with other inputs. For example, the marginal value of irrigation water for crop production is the value of the incremental crop yield that can be attributed to irrigation, rather than to labor, fertilizer, and other inputs.
Natural capital	Here we focus on living, renewable forms of natural capital, which constitute a stock – of an ecosystem and the biota that makes it up – that generates a flow of ecosystem services.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

H. Tallis (✉) • A. Guerry
The Natural Capital Project, Department of Biology and Woods Institute,
Stanford University, Stanford, CA 94305-5020, USA
e-mail: htallis@stanford.edu; anne.guerry@stanford.edu

G.C. Daily
The Natural Capital Project, Department of Biology and Woods Institute,
Stanford University, Stanford, CA 94305-5020, USA

Department of Biological Sciences, Stanford University, Y2E2 Building,
473 Via Ortega, Room 227C, Stanford, CA 94305-5020, USA
e-mail: gdaily@stanford.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_6,

> For example, a forest constitutes a stock that generates a flow of timber, carbon sequestration, water quality, biodiversity, serenity, and other benefits, depending upon how it is managed. (Fossil fuels and other minerals constitute nonliving natural capital, which is generally nonrenewable on time scales of interest to society.)

Definition

Ecosystem services are essential to sustaining and fulfilling human life, and yet their supply is seriously threatened by the intensification of human impacts on the environment. Over the past decade, efforts to value and protect ecosystem services have been promoted by many as the last best hope for making conservation mainstream – attractive and commonplace worldwide. In theory, if institutions recognize the values of nature, then we can greatly enhance investments in conservation and foster human well-being at the same time. In practice, scientific and policy communities have not yet developed the scientific basis or the policy and finance mechanisms for integrating natural capital into resource and land-use decisions on a large scale.

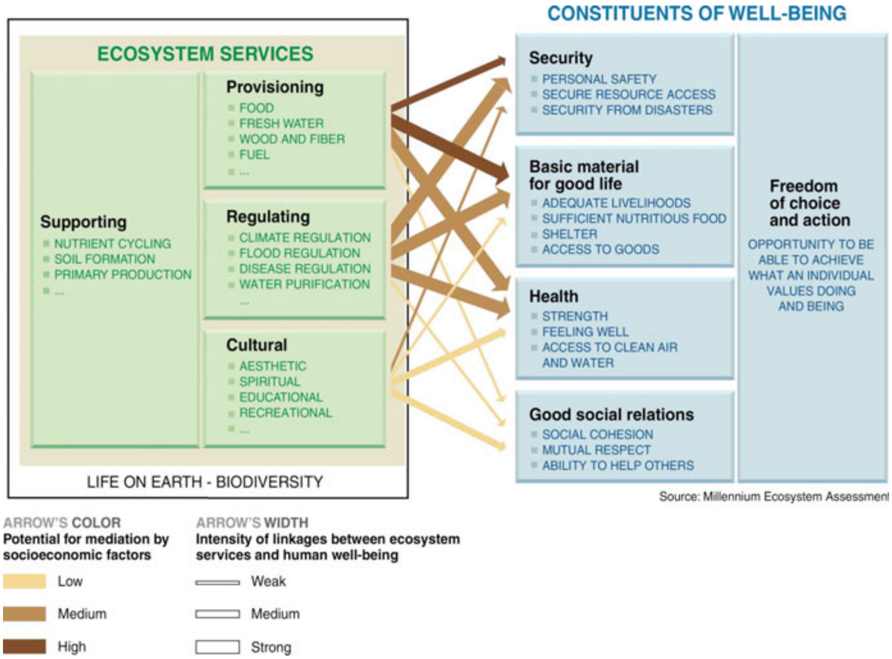
Introduction

This entry provides an overview of issues concerning the identification, biophysical and economic characterization, and safeguarding of ecosystem services. The concept of ecosystem services has a long written history, reaching back at least as far as Plato. We review this history, including a focus on recent advances such as the Millennium Ecosystem Assessment, and advances in spatial modeling, economic valuation, and policy implementation. We provide examples of novel finance and policy mechanisms, including water funds, marine spatial planning, land-use planning and human development, and global policy efforts. We conclude with a discussion of the largest research and implementation challenges in this field, identifying the issues that will frame the future growth of the concept of ecosystem services.

What are Ecosystem Services?

Definition and Classification

Ecosystem services are defined simply as the benefits that people obtain from ecosystems [38]. They sustain and fulfill human life and flow from many conditions and processes of ecosystems, and the species making them up [14]. The processes



Source: Millennium Ecosystem Assessment

Fig. 6.1 Ecosystem service categories and their linkages to human well-being as described in the Millennium Ecosystem Assessment [38]

and features generating ecosystem services (ES) are so tightly interconnected that any classification is inherently somewhat arbitrary. The most widely used classification was developed through the Millennium Ecosystem Assessment (MA) and identifies four classes of ES based on their types of benefits to society:

1. *Provisioning services* including the production of goods such as food, water, timber, and fiber
2. *Regulating services* that stabilize climate, moderate risk of flooding and disease, and protect or enhance water quality
3. *Cultural services* that provide recreational, aesthetic, educational, community, and spiritual opportunities
4. *Supporting services* that underlie provision of the other three classes of benefits, including soil formation, photosynthesis, nutrient cycling, and the preservation of options (Fig. 6.1; [38]).

The classification of ES is still a topic of debate and several other classification approaches have been suggested [7, 19, 22, 67].

System- and scale-neutral, the ecosystem services framework applies equally to terrestrial, freshwater, and marine ecosystems and their processes, anywhere on the spectrum from relatively pristine to heavily managed conditions. Indeed, all ecosystems provide, to differing degrees, a set of ES. Human conversion of ecosystems from one type to another is often motivated by a desire for a different set of ES, though consideration of the services of the two systems and their tradeoffs is often incomplete.

Ecosystem Services Across Systems

A wide range of ES is generated in the terrestrial realm, by croplands, natural and managed forests, grasslands, and wetlands. In each of these systems, for example, vegetation can protect and enhance soils, preventing their loss through erosion and improving fertility by retaining moisture and storing and recycling nutrients. Vegetation and soils together regulate the quantity, quality, and timing of water flows, thus moderating floods and droughts and providing cleaner, more reliable supplies [8].

Forests stand out as important in regulating water and carbon cycles [30]; in their strong influence on local, regional, and global climate [30, 49]; and because of the multiple, interacting threats to their future (e.g., [44, 66]). They also provide natural products for subsistence use or sale including timber, firewood, mushrooms, fruits and seeds, medicinal plants, rubber, cork, and bushmeat. Forest and woodland habitats harbor species that provide pollination and pest control to commercial or subsistence crops. Grassland and other dryland systems play these same critical roles in addition to supporting vast livestock populations [38]. Wetlands occupy a small fraction of Earth's surface, but dominate the landscape where they are concentrated and provide a wide array of water quality, flood mitigation, coastal protection, and biogeochemical services [38]. Each of these systems, however natural or managed, can provide habitat for biodiversity and opportunities for recreational activities, spiritual experiences, and creative, cultural expression.

Freshwater ecosystems provide a suite of highly visible and widely appreciated ES [53]. The freshwater regulated by terrestrial systems and the atmosphere is used for drinking, hydropower production, irrigation, household activities (washing, etc.), industrial purposes (cooling, manufacturing, etc.), and cultural experiences. People also gain large revenues and nutrition from freshwater fisheries and aquaculture. Less appreciated is the value of sediment transport and deposition in rivers that supply river reaches and downstream beaches with important sand and gravel resources. Wetlands and other aquatic vegetation can regulate flood waters and cycle nutrients, improving water quality. Finally, freshwater systems serve as pathways for human transportation and recreational or cultural activities.

Marine ecosystems also provide all four classes of ecosystem services described in the MA. Marine fisheries and aquaculture provide nutrition, feed for animals, livelihoods, and important recreational and cultural opportunities. Harvests of other species for food additives, cosmetics, and pharmaceuticals also support health, nutrition, and livelihoods. Marine biogenic habitats (such as coral reefs, oyster reefs, and kelp forests) regulate natural hazards including storm surges, and may play a critical role in helping coastal communities adapt to sea level rise. Marine systems also transform, detoxify, and sequester wastes.

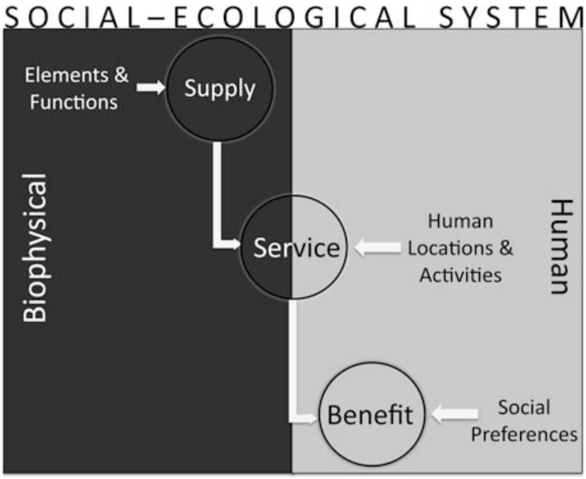


Fig. 6.2 Three measurement points for ecosystem services [63]. Supply metrics deal only with the biophysical system underpinning the service of interest. Service metrics include critical information linking supply to beneficiaries. Benefit metrics weigh the level of service based on people's preferences or social policy goals

In addition, oceans are the center of the global water cycle; they hold 96.5% of the Earth's water [24] and are a primary driver of the atmosphere's temperature, moisture content, and stability [12]. Oceans are also key players in the global cycles of carbon, nitrogen, oxygen, phosphorus, sulfur, and other major elements [51] and are responsible for approximately 40% of global net primary productivity [39, 61]. Finally, coastal communities reap many benefits from coastal tourism (one the one of world's most profitable industries [65]), and numerous coastal communities define their very identities in relation to the sea and all it brings.

The Ecosystem Service Supply Chain

Ecosystem services flow to people along a supply chain from biophysical systems to people [63]. All services are generated by some function or element of a natural or managed system (Fig. 6.2). The full suite of these elements or functions can best be considered in three discrete steps: supply, service, and benefit. For example, consider protection from coastal storm surges. Many different types of coastal elements (e.g., coral reefs, mangroves, oyster beds, barrier islands) confer protection from storm surges by attenuating waves. The full set of locations of these coastal elements represents the supply of protection from storms. People do not receive storm surge protection from all of these locations, however, because some are far from human infrastructure and settlement. Both the distribution of human infrastructure and settlement, together with the location and condition of supply, are required to give a clear picture of how much "service" is actually delivered at a given time.

Finally, the service delivered to human communities is often valued differently, depending on the context. For example, coastal protection services provided by nearshore habitats to easily accessible, popular, public beaches might be seen as more valuable, or providing greater benefit, than those to more remote sites.

History of the Concept of Ecosystem Services

It is primarily through disruption and loss that the nature and value of ecosystem services has been illuminated. For instance, deforestation has demonstrated the critical role of forests in the hydrological cycle – in particular, in mitigating floods, droughts, the erosive forces of wind and rain, and the silting of dams and irrigation canals. Release of toxic substances, whether accidental or deliberate, has revealed the nature and value of physical and chemical processes, governed in part by a diversity of microorganisms, that disperse and break down hazardous materials. Thinning of the stratospheric ozone layer sharpened awareness of the value of its service in screening out harmful ultraviolet radiation. And the loss of coastal wetlands has brought into relief their importance in regulating coastal hazards such as hurricanes and tsunamis.

Initial Development of the Ecosystem Services Concept

A cognizance of ecosystem services, expressed in terms of their loss, dates back at least to Plato and probably much earlier:

> What now remains of the formerly rich land is like the skeleton of a sick man with all the fat and soft earth having wasted away and only the bare framework remaining. Formerly, many of the mountains were arable. The plains that were full of rich soil are now marshes. Hills that were once covered with forests and produced abundant pasture now produce only food for bees. Once the land was enriched by yearly rains, which were not lost, as they are now, by flowing from the bare land into the sea. The soil was deep, it absorbed and kept the water. . ., and the water that soaked into the hills fed springs and running streams everywhere. Now the abandoned shrines at spots where formerly there were springs attest that our description of the land is true. (Plato)

Mooney and Ehrlich [40] trace modern concern for ecosystem services to George Perkins Marsh, a lawyer, politician, and scholar. Indeed, his 1864 book *Man and Nature* describes a wide array of services, again, often expressed in terms of their loss. Remarking on the terrain of the former Roman Empire, he notes that it "is either deserted by civilized man and surrendered to hopeless desolation, or at least greatly reduced in both productiveness and population" (p. 9). He continues, describing the reduction of hydrological services: "Vast forests have disappeared from mountain spurs and ridges, the vegetable earth . . . [is] washed away; meadows, once fertilized by irrigation, are waste and unproductive, because . . . the springs that fed them dried up; rivers famous in history and song have shrunk to humble brooklets" (p. 9). He also draws connections between deforestation and climate: "With the disappearance of the forest, all is changed. At one season, the Earth parts with its warmth by radiation to an open sky – receives, at another, an immoderate heat from the unobstructed rays of the sun. Hence the climate becomes excessive, and the soil is alternately parched by the fervors of summer, and seared by the rigors of winter. Bleak winds sweep unresisted over its surface, drift away the snow that sheltered it from the frost, and dry up its scanty moisture" (p. 186). Finally, he even wrote of decomposition services: "The carnivorous, and often the herbivorous insects render an important service to man by consuming dead and decaying animal and vegetable matter, the decomposition of which would otherwise fill the air with effluvia noxious to health" (p. 95).

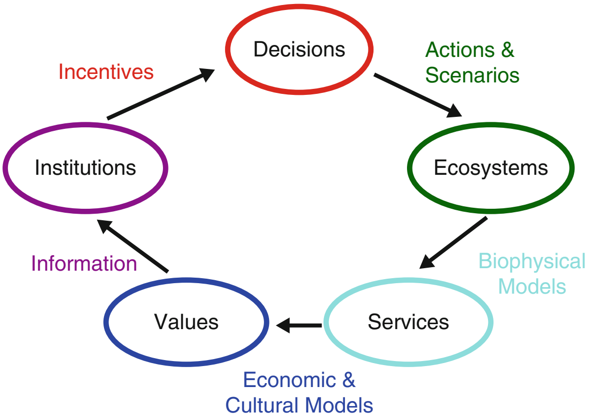
Other eloquent writers on the environment emerged following World War II, including Fairfield Osborn (*Our Plundered Planet*, 1948), William Vogt (*Road to Survival*, 1948), and Aldo Leopold (*A Sand County Almanac and Sketches from Here and There*, 1949). Each discusses ecosystem services without using the term explicitly. In *The Population Bomb* (1968), Paul Ehrlich describes anthropogenic disruption of ecosystems and the societal consequences of doing so, addressing the need to maintain important aspects of ecosystem functioning. Along these lines, the *Study of Critical Environmental Problems* (1970) presents a list of key "environmental services" that would decline with a decline in "ecosystem function." This list was expanded upon by Holdren and Ehrlich [29]. Meanwhile, in the 1960s and 1970s, economists set out to measure "the value of services that natural areas provide" ([35], p. 12), with efforts focused on agricultural production [3], renewable resources [11, 34], nonrenewable resources [18], and environmental amenities [23].

By the early 1980s, efforts were initiated to investigate two questions: the extent to which ecosystem function (and the delivery of services) depends on biodiversity, and the extent to which technological substitutes could replace ecosystem services. The first question is addressed in chapter Species Diversity Within and Among Ecosystems, this volume. The second question was tackled by Ehrlich and Mooney [21]. Work on these topics proliferated and, in 1997, a collective effort was made to synthesize the wealth of scientific information that had accumulated on the functioning of ecosystem services, with a preliminary exploration of their economic value, and of key issues meriting further work [14].

Recent Advances

Four major advances of the last decade have revitalized research on ecosystem services and brought them into the public eye. First, the MA represented a visionary and seminal step in global science – it was the first comprehensive global assessment of the status and trends of all of the world's major ecosystem services. It was requested by United Nations Secretary General Kofi Annan in 2000 and carried out

Fig. 6.3 A framework showing key elements for integrating ES into decision-making [16]. One could link any two ovals, in any direction, in different decision contexts



between 2001 and 2005 with contributions from over 1,360 experts worldwide. The key finding of this assessment was that two thirds of the world's ecosystem services were declining [38]. This captured the attention of world leaders and emphasized the connections between human decisions and the natural environment that feed back to the human condition via changes in the flow of ecosystem services.

Work following the MA clarified this chain of connections (Fig. 6.3) [16]. Human decisions shape individuals' actions relating to the use of land, water, oceans, and other elements of natural capital. These actions often alter the state or functioning of ecosystems, which in turn provide altered flows of benefits (goods or services) to people. People express different values (monetary, cultural) associated with these altered streams of benefits and it is the expression of these values that leads to changes in institutions that guide decisions. The following three recent advances all concern the connections in this flow.

A suite of recent advances has greatly improved understanding of the links between ecosystem functions and processes and the provision of ecosystem services (Fig. 6.3). For some ecosystem services, we now better understand the key ecological system components that drive provision (e.g., [33]) and we can now measure (e.g., [56]) and model, with uncertainty, the impacts of land use and resource management decisions on a wider variety of ecosystem processes and associated services. Ecological science has also advanced spatially explicit modeling, which is essential for mapping ecosystem services and their flows to people (e.g., [10, 27, 43, 57]). Finally, we are starting to see patterns in how multiple ecosystem services and biodiversity change in relation to each other. Recent work has started characterizing bundles of ecosystem services, and exploring their synergies and trade-offs (e.g., [4, 6, 20, 42, 43]).

Further, economic valuation methods have been applied to the spatial provision of ecosystem services to estimate the monetary value of benefits and, in some cases, the distribution of those benefits to various segments of society [42, 45, 52, 64]. In addition, qualitative and quantitative methods from other social sciences have been applied to gain better understanding of the social and cultural importance of ecosystem services (e.g., [38]).

Lastly, experiments in payments for ecosystem services [47, 48, 69], in ecosystem-based management [2], and in regional planning have begun, giving us opportunities to learn about how science can play a role in altering institutions, and how institutions alter decisions and the resulting flow of ecosystem services. The following section describes some of these efforts in more detail.

Incorporating Ecosystem Services into Decisions

Today, the urgent challenge is to move from theory to practical implementation of ES tools and approaches in resource decisions taken by individuals, communities, corporations, and governments. The framework in Fig. 6.3 connects the science of quantifying services with valuation and policy to devise payment schemes and management actions that take account of ecosystem services. This connection is expressed in the real world in a variety of ways across scales from local to global.

A great number and diversity of efforts to implement the ES framework have emerged worldwide over the past decade. Individually, most of these efforts are small and idiosyncratic. But collectively, they represent a powerful shift in the focus of conservation organizations and governments (primarily) toward a more inclusive, integrated, and effective set of strategies [15]. Taken together, these efforts span the globe and target a full suite of ecosystem services, including principally forest-generated services of carbon sequestration, water supply, flood control, biodiversity conservation, and enhancement of scenic beauty (and associated recreation/tourism values) [26, 62].

Many local or regional ES efforts focus on a single service that stands out as sufficiently important, from economic and political perspectives, to overcome the activation energy required to protect it. Under the institutional umbrella created for the focal service, it is possible that other services may be at least partially protected. Beginning in the late 1990s, larger-scale investment in natural capital for water flow regulation in China – and for a broad suite of ES in Costa Rica – set pioneering examples that are now being adapted elsewhere and scaled up.

Next, we briefly describe some contrasting models of success, at different scales and in different kinds of social-ecological systems. In each case, there is an acute or looming crisis, innovative leadership, and pursuit of dual goals: improving both human and ES condition.

Local Scale: Water Funds

New York City made one of the first and most famous investments in ecosystem service provision in recent history, in the mid-1990s. The city invested ca. USD1.5 billion in a variety of watershed protection activities to improve drinking water quality for 10 million users rather than spending the estimated USD6-8 billion needed (excluding annual operating and maintenance costs) for building a new filtration plant. This seminal example is widely cited as evidence of the business case for investing in natural capital instead of built capital [15]. Yet the effort remains very much an experiment in the science and policy of investing in natural capital, and one on which there is international focus.

Globally, watersheds are now emerging as the target of a range of creative policy and finance mechanisms that link beneficiaries to suppliers through a payment system. In these “water funds,” water users voluntarily pay into a pool that is collectively managed by contributors and invested in watershed management improvements. The Nature Conservancy (TNC) has now established more than ten water funds in Latin America, has plans to create 22 more by 2015 [25], and is exploring the possibility of establishing some of the first funds in Africa.

Agua por la Vida y la Sostenibilidad, one of the recently established water funds, demonstrates the diversity of water users that are becoming engaged in these funds and the kinds of watershed management changes these funds motivate. Formally established in the Cauca Valley, Colombia in 2009, this water fund is supported by the region’s sugarcane grower’s association (PROCAÑA), the sugar producers’ association (ASOCAÑA), 11 local watershed management groups, TNC and a Colombian peace and justice nongovernment organization (Vallenpaz). Each member of the water fund voluntarily pays a self-determined amount into the fund that is then jointly managed by the members to improve landscape management in 11 watersheds covering over 3,900 km^2.

Members in this fund have currently committed to contributing USD10 million over 5 years to be invested in five kinds of management changes: protection of native vegetation, restoration of denuded lands, enrichment of degraded forests, fencing of rangelands, and implementation of silvopastoral practices. The fund is starting a monitoring program that will ensure that these investments lead to measurable improvements in water quality for approximately one million water users downstream and significant improvements in terrestrial and freshwater biodiversity.

Local Scale: Coastal and Marine Spatial Planning

People commonly think of oceans as relatively featureless expanses that defy the drawing of lines on maps. However, recent political and scientific advances have highlighted the need for a comprehensive approach to planning marine and coastal uses and the need for practical tools to make this more comprehensive approach a reality on the ground and in the water. In a marine spatial plan, a wide range of uses of the marine environment are put on one map. But an understanding of how such plans are likely to yield changes in the delivery of the broad range of services people receive from the system has, until recently, remained elusive.

Along the west coast of Vancouver Island Canada, multiple, often competing interests are struggling to define the future character of the place. Existing extractive, industrial, and commercial uses; traditional First Nations subsistence and ceremonial uses; recreation and tourism; and emerging ocean uses such as the extraction of wave energy are all in the mix. The West Coast Aquatic Management Board (WCA) is charged with creating a marine spatial plan for the region. WCA is a public-private partnership with participation from four levels of government (Federal, Provincial, local, and First Nations), and diverse stakeholders. Ultimately, WCA's vision is to manage resources for the benefit of current and future generations of people and nonhuman species and communities.

Some key pillars of the partnership's strategy are to: use a precautionary, ecosystem-based approach to protect, maintain, and restore marine and coastal resources; respect and protect First Nations' food, social and ceremonial requirements and treaty obligations; integrate expertise and knowledge from First Nations, local, scientific, and other sources; ensure broad participation in the planning process; and foster initiatives that maintain or enhance opportunities for coastal communities to benefit from local resources, while achieving sustainable social, cultural, and economic benefits for the region. WCA has partnered with the Natural Capital Project to explore how alternative spatial plans might affect a wide range of ES and to provide information about trade-offs among ES.

Key considerations for WCA and their stakeholders include balancing important industrial and commercial activities (such as shipping, mining, logging, aquaculture, and fisheries), increased development of tourism and recreation, renewable energy generation, and a strong cultural desire for sustaining the remote, wild feeling of the place. WCA is exploring the suitability of alternative regions for these different activities. For example, maps of coastal vulnerability to erosion and flooding from storm surge are helping to direct coastal development permits to low-risk areas. Similar maps of the value of captured wave energy are being overlaid with existing ocean uses (e.g., fishing and recreational activities) to highlight regions of high wave energy value, where wave energy generation facilities might be constructed while having minimal impacts on other activities. Examinations of trade-offs among aquaculture (finfish, shellfish), wild salmon fisheries, recreation (e.g., kayaking, whale watching, and diving), coastal development (on the coast, as well as floathomes), and habitat and water quality are underway.

The general framework of ES and ES modeling, in particular, is helping to articulate connections between human activities that are often considered in isolation, to align diverse stakeholders around common goals, and to make implicit decisions explicit. ES modeling results have informed early iterations of the marine spatial plan and will inform the creation of the final plan in 2012.

National Scale: Land-Use Planning and Human Development in China

The ecosystem service investments being made in China today are impressive in their goals, scale, duration, and innovation. Following massive droughts and flooding in 1997–1998, China implemented several national forestry and conservation initiatives, into which investments exceeded 700 billion yuan (ca. USD100 billion) over 2000–2010 [37, 70]. The larger and older of these initiatives are being rigorously evaluated to determine their biophysical and socio-economic impacts, to improve their design and efficacy.

These initiatives have dual goals: to secure critical natural capital through targeted investments across landscapes and regions, and to alleviate poverty through targeted wealth transfers from coastal provinces to inland regions where many ES originate. The Chinese government aims to reduce the loss of soil, improve water retention, reduce desertification, and generally protect biodiversity and ecosystems in the west of the country for flood control, hydropower production efficiency, irrigation supply, more productive agriculture, and ecotourism. In addition, it wants to change the economic structure in rural areas to increase local household income while simultaneously making local households' patterns of land utilization and agricultural production more sustainable [36, 37].

The initiatives include two national PES programs, the Natural Forest Conservation Program (NFCP) and the Sloping Land Conversion Program (SLCP), established in 1998 and 1999 respectively. Implementation was tested in a few provinces, and then rapidly scaled to the whole country. Evaluation of the programs shows significant achievement of the biophysical goals, with remarkably rapid land conversion in the desired directions. For example, by the end of 2006, the SLCP had converted ca. 9 million hectares of cropland into forest/grassland and had afforested ca. 12 million hectares of barren land. Village-level field measurements have shown not only that the payments for ES have altered land use patterns, but in turn soil erosion has been decreased in some areas by as much as 68% [9].

Overall social impacts of the programs are mixed. In some places, payment levels and types are leading to improvements in economic measures of well-being, whereas in others payments were not sufficient to compensate for loss of income from shifting livelihoods [37]. In addition, in some places where participation in the SLCP has significant positive impacts upon household income, it has not yet transferred labor toward non-farming activities as the government wished [36]. Payments are now being adjusted to improve success in achieving goals of poverty alleviation and growth of new economic sectors in rural areas.

China is also now establishing a new network of Ecosystem Function Conservation Areas (EFCAs), specifically for ES provision. Their exact delineation is now being determined through quantitative ecosystem service mapping and valuation. They are expected to span ca. 25% of the country.

The current and potential future impacts of ES investments in China are enormous, certainly within the country – and also globally, in the form of enhanced carbon sequestration and reduced dust export, and perhaps most importantly in lessons on making the investments needed in natural capital and human well-being everywhere.

International Scale: Global Policy and Research Efforts

As described above, the MA was the first major effort to establish ES in the international policy arena. Activities stemming from that effort are now aimed at bringing countries together in making tangible commitments to safeguard ES (e.g., 2020 targets for the Convention on Biodiversity) and to assess national and international progress toward those commitments (e.g., through Group on Earth Observations Biodiversity Observation Network (GEO BON) and the Programme on Ecosystem Change and Society (PECS), which synthesize knowledge for the International Platform for Biodiversity and Ecosystem Services (IPBES), formally established in 2010). Several new international research efforts aim to feed into these international processes, including the Natural Capital Project, The Resilience Alliance, and the Stockholm Resilience Centre. Other entities are focused on establishing and tracking ES markets, as a mechanism for bringing larger attention to ES benefits to society (e.g., The Katoomba Group and The Ecosystem Marketplace, both initiated by Forest Trends). As an example of many burgeoning international efforts, we describe in greater detail the Natural Capital Project.

The Natural Capital Project (NatCap) (www.naturalcapitalproject.org) is an international partnership working to align economic forces with conservation, by developing tools that make incorporating natural capital into decisions easy and replicable; by demonstrating the power of these tools in important, contrasting places; and by engaging leaders globally. NatCap is developing InVEST, a family of tools for Integrated Valuation of Ecosystem Services and Tradeoffs.

InVEST helps decision makers visualize the impacts of potential policies by modeling and mapping the delivery, distribution, and economic value of ES under alternative scenarios (for more information, see [30]). The outputs identify trade-offs and compatibilities between environmental, economic, and social benefits. InVEST is designed for use as part of an active decision-making process (Fig. 6.4) and can be applied at local, regional, or global scales. The first phase of the approach involves working with stakeholders to identify critical management decisions and to develop scenarios that project how the provision of services might change in response to those decisions as well as to changing climate, population, etc. Based on these scenarios, a modular set of models quantifies and maps ES. The outputs of these models provide decision makers with information about costs, benefits, trade-offs, and synergies of alternative investments in ES provision.

NatCap is using InVEST in major natural resource decisions in diverse contexts around the world, including in the three examples given above (water funds, coastal and marine spatial planning, and land-use planning and human development in China). The aim is to demonstrate the power of these approaches and to learn how to replicate and scale up models of success. The Project is engaged in a suite of international efforts, including GEO BON and IPBES, to offer a common, unifying platform for regional and national efforts that are spawned by these initiatives.

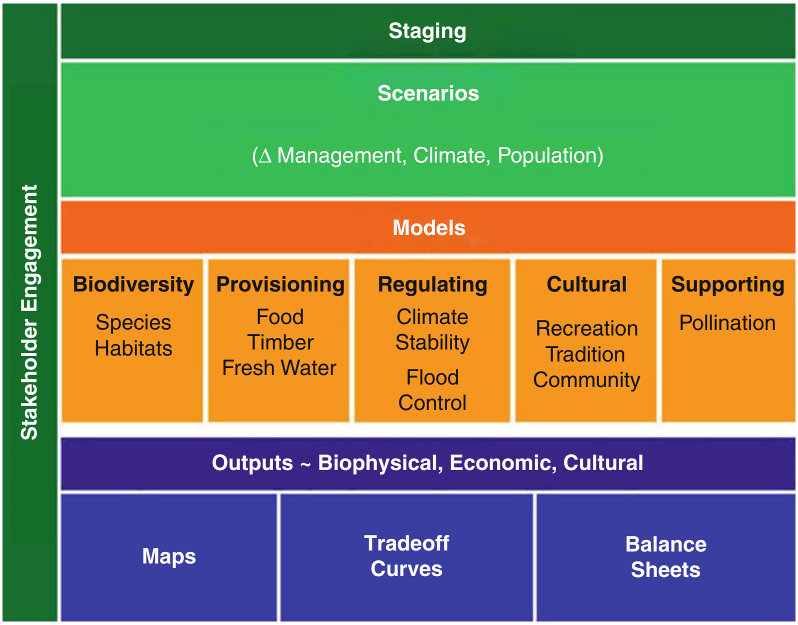


Fig. 6.4 An iterative process for integrating ecosystem services into decisions. The process begins with stakeholder engagement around impending decisions, with a focus on realistic, alternative scenarios for the future. The modeling is shaped by stakeholders, and typically focused on the services and scenarios deemed most important. Outputs are displayed in accordance with stakeholder preferences, in the form of maps, trade-off curves, and/or balance sheets. These can be expressed in biophysical (e.g., tons of carbon), economic (e.g., dollars), or cultural (e.g., visitor-days) terms

Future Directions

With the rapid rate of development of ES mapping, from the biophysical and economic modeling through to policy application in diverse socioeconomic contexts, it is likely that great advances will be made in coming years. What we report here is only a beginning. There are key arenas in which further learning is crucial to understand what drives variation in the provision of ES, how they percolate through various arms of society, and how social reaction leads to sustainable or unsustainable change in ES provision.

Relating Ecosystem Services and Human Health

The relationships between biophysical attributes of ecosystems and human health are complex [41]. Destruction of natural ecosystems can at times improve aspects of public health. Draining swamps, for example, can reduce habitat for the mosquito vector that transmits the parasite that causes malaria. On the other hand, destruction of other systems can have sharp negative consequences for human health. There is emerging evidence that loss of tropical forests, for example, leads to an increase in transmission of malaria [1, 50]. Similarly, fragmentation of, and biodiversity loss from, eastern North American forests is associated with an increase in lyme disease [32].

Natural and managed ecosystems provide many services that sustain human health, through provision of human nutrition (especially of protein and micronutrients); purification and regulation of drinking water; regulation of air quality; regulation of vector-borne disease; and psychological benefits. There is a great need for research illuminating the links between biodiversity, ecosystem conditions and processes, and human health.

Trade-offs and Synergies

The relationship between ES and biodiversity and among different ES varies with socio-ecological context. In some cases, clear trade-offs and synergies among services have been defined in specific contexts [54], but there is still much to be learned about what determines the nature of these relationships. Advancing this knowledge is essential because policies addressing management change can only be successful if management controls ES relationships. If policies are established to align multiple ecosystem services, but biophysical conditions in the system lead to innate trade-offs among services, management changes are bound to fail in delivering the desired improvements to social benefits.

Distributional Effects

Much of the science of mapping ES has focused on identifying where ES are generated and where they are delivered. However, less work has focused on identifying to whom ES actually flow. This connection is essential if policies addressing ES delivery are to be equitable and either improve the well-being of the poor or avoid unintended distributional consequences. Past work in this arena has focused on overlaying maps of ES provision with an array of poverty indicators (e.g., [68]). Missing from this spatial analysis is information on access to and ability to control the delivery of ES. In many cases (e.g., for services such as clean drinking water, hydropower production, agriculture, water for irrigation, wave power generation), the actual delivery of services to specific people is affected by the location of infrastructure or institutions regulating access to resources. New science is needed that allows the ready mapping of these connections and the prediction of how they will change under future conditions.

Dynamic Effects: Shocks and Uncertainty

Dynamic changes, such as in climate and in the nitrogen cycle – as well as changes arising through economic development and evolving human preferences over time – are very important. The possibility of feedbacks within ecosystems, and between ES and human behavior, is a key area for further development. Feedback effects can give rise to thresholds and rapid changes in systems that can fundamentally alter system outcomes [60]. The ability to incorporate shocks and the possibility of surprises is another area where further development is needed. Fires, droughts, and disease all can have major influences on ecosystems and affect the services produced. Changes in economic conditions or fads in human behavior can similarly cause major changes in systems (e.g., financial crises). The occurrence of each of these and other potential disturbances is difficult to predict but virtually certain to come about. Understanding their likely impacts on ecological and social systems will help us prepare for them.

Valuation in Monetary and Non-monetary Terms for Decision-Making

Monetary valuation of ES is not nearly as prevalent as sometimes assumed. More typically, real-world applications of the ES framework rely on biophysical values to inform policy design, such as measures of water quality or flood risk.

Value is not always easily characterized or fully captured in monetary terms, so it is important to characterize value in multiple dimensions, including health, livelihood support, cultural significance, etc. (e.g., [17]). This will help ensure that valuation and broader decision-making approaches are inclusive of the range of benefits and people concerned [28]. Interdisciplinary efforts are presently underway to create a conceptual framework that is useful both in theory and in practice for a broad suite of cultural ES.

Institutional Design

However ES are measured, there is a need for political and social science research to design institutions and policy mechanisms that better capture externalities. Efforts such as national accounts are blossoming now, but it is unclear how they will evolve and how successful governments will be at incorporating natural capital into national measures of wealth. There is great work to be done in determining the merits and limitations of alternative policy and finance mechanisms, in different economic, governance, and other social contexts (e.g., [5, 46, 58, 59]). There is also great work to be done in developing institutions that achieve representation and participation by stakeholders as part of adaptive governance systems (e.g., [13, 55]).

Conclusions

Ecosystem services have had a relatively long history through indirect recognition of the importance of nature for the persistence of the human endeavor. There are scientific challenges for biogeochemists, hydrologists, ecologists, economists, anthropologists, and other social scientists to understand how human actions affect ecosystems, the provision of ES, and the value of those services. At least as demanding are the social and political questions associated with incorporating this understanding into decision-making. There is also a need to design effective and enduring institutions to manage, monitor, and provide incentives that reflect the social values of ecosystem services. Information is becoming more readily available for individuals, corporate managers, and government officials who make decisions affecting ecosystems and the services to consider a more complete set of costs and benefits associated with their choices. We are likely to see continuing growth in our scientific ability to measure and predict changes in ES, our ability to design policies and institutions that accurately represent these changes and in turn, the ability of the environment to continue providing the many benefits society needs to prosper.

Bibliography

Primary Literature

1. Afrane YA, Little TJ, Lawson BW, Githeko AK, Yan G (2008) Deforestation and vectorial capacity of *Anophiles gambiae* Giles mosquito in malaria transmission, Kenya. Emer Infect Dis 14:1533–1538
2. Barbier E, Koch E, Silliman B et al (2008) Coastal ecosystem-based management with non-linear ecological functions and values. Science 319:321–323
3. Beattie B, Taylor CR (1985) The economics of production. Wiley, New York
4. Bennett E, Peterson G, Gordon L (2009) Understanding relationships among multiple ecosystem services. Ecol Lett 12:1394–1404
5. Berkes F, Colding J, Folke C (2003) Navigating social-ecological systems. Cambridge University Press, Cambridge
6. Boody G, Vondracek B, Andow D et al (2005) Multifunctional agriculture in the United States. Bioscience 55:27–38
7. Boyd JW, Banzhaf HS (2005) Ecosystem services and government accountability: the need for a new way of judging Nature's value. Resources 158:16–19, Summer
8. Brauman KA, Daily GC, Duarte TK, Mooney HA (2007) The nature and value of ecosystem services: an overview highlighting hydrologic services. Annu Rev Environ Resour 32:67–98
9. Cao S, Zhong B, Yue H, Zeng H, Zeng J (2009) Development and testing of a sustainable environmental restoration policy on eradicating the poverty trap in China's Changting County. Proc Natl Acad Sci 106:10712–10716
10. Chan K, Shaw R, Cameron D et al (2006) Conservation planning for ecosystem services. Public Libr Sci Biol 4:2138–2152
11. Clark C (1990) Mathematical bioeconomics, 2nd edn. Wiley, New York

12. Colling A (2001) Ocean circulation. Butterworth Heineman/The Open University, Milton Keynes
13. Cowling R, Egoh B, Knight AT et al (2008) An operational model for mainstreaming ecosystem services for implementation. Proc Natl Acad Sci 105:9483–9488
14. Daily GC (ed) (1997) Nature's services. Island Press, Washington, DC
15. Daily GC, Ellison K (2002) The new economy of nature. Island Press, Washington, DC
16. Daily GC, Polasky S, Goldstein J et al (2009) Ecosystem services in decision-making: time to deliver. Front Ecol Environ 7:21–28
17. Dasgupta P (2001) Human well-being and the natural environment. Oxford University Press, Oxford
18. Dasgupta P, Heal G (1979) Economic theory and exhaustible resources. Cambridge University Press, Cambridge, UK
19. De Groot D, Wilson MA, Boumans RMJ (2002) A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol Econ 41:393–408
20. Egoh B, Reyers B, Rouget M et al (2008) Mapping ecosystem services for planning and management. Agric Ecosyst Environ 127:135–140
21. Ehrlich PR, Mooney HA (1983) Extinction, substitution, and ecosystem services. BioScience 33:248–254
22. Fisher B, Turner RK, Morling P (2008) Defining and classifying ecosystem services for decision making. Ecol Econ 68:643–653
23. Freeman AM III (1993) The measurement of environmental and resource values. Resources for the Future, Washington, DC
24. Gleick P (1996) Water resources. In: Schneider S (ed) Encyclopedia of climate and weather. Oxford University Press, New York, pp 817–823
25. Goldman R, Benítez S, Calvache A, Rámos A (2010) Water funds: protecting watersheds for nature and people. The Nature Conservancy, Arlington
26. Goldman RL, Tallis H, Kareiva P, Daily GC (2008) Field evidence that ecosystem service projects support biodiversity and diversify options. Proc Natl Acad Sci 105:9445–9448
27. Harrison J, Bouwman A, Mayorga E, Seitzinger S (2010) Magnitudes and sources of dissolved inorganic phosphorus inputs to surface fresh waters and the coastal zone: a new global model. Glob Biogeochem Cycles 24:271–279
28. Heal G (2000) Nature and the marketplace: capturing the value of ecosystem services. Island Press, Washington, DC
29. Holdren J, Ehrlich P (1974) Human population and the global environment. Am Sci 62:282–292
30. Jackson RB, Randerson JT, Canadell JG, Anderson RG, Avissar R, Baldocchi DD, Bonan GB, Caldeira K, Diffenbaugh NS, Field CB, Hungate BA, Jobbágy EG, Kueppers LM, Nosetto MD, Pataki DE (2008) Protecting climate with forests. Environ Res Lett 3:1–5
31. Kareiva PK, Tallis H, Ricketts TH, Daily GC, Polasky S (eds) (2011) Natural capital: theory & practice of mapping ecosystem services. Oxford University Press, Oxford
32. Keesing F et al (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature 468:647–652
33. Kremen C (2005) Managing for ecosystem services: what do we need to know about their ecology? Ecol Lett 8:468–479
34. Krutilla JV (1967) Conservation reconsidered. Am Econ Rev 57:777–786
35. Krutilla JV, Fisher AC (1975) The economics of natural environments: studies in the valuation of commodity and amenity resources. Johns Hopkins University Press, Baltimore
36. Li J, Feldman M, Li S, Daily GC (2011) Rural household income and inequality under payment for ecosystem services: the Sloping Land Conversion Program in Western China. Proc Natl Acad Sci 108:7721–7725
37. Liu J, Li S, Ouyang Z et al (2008) Ecological and socioeconomic effects of China's policies for ecosystem services. Proc Natl Acad Sci 105:9489–9494

38. MA (Millennium Ecosystem Assessment) (2005) Ecosystems and human well-being: the assessment series (four volumes and summary). Island Press, Washington, DC
39. Melillo JM, Mcguire AD, Kicklighter DW, Moore B, Vorosmarty CJ, Schloss AL (1993) Global climate-change and terrestrial net primary production. Nature 363:234–240
40. Mooney HA, Ehrlich PR (1997) Ecosystem services: a fragmentary history. In: Daily GC (ed) Nature's services. Island Press, Washington, DC, pp 11–19
41. Myers S, Patz J (2009) Emerging threats to human health from global environmental change. Annu Rev Environ Resour 34:223–252
42. Naidoo R, Ricketts TH (2006) Mapping the economic costs and benefits of conservation. PLoS Biol 4:e360
43. Nelson E, Mendoza G, Regetz J et al (2009) Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. Front Ecol Environ 7:4–11
44. Nepstad D, Carvalho G, Barros AC, Alencar A, Capobianco JP, Bishop J, Moutinho P, Lefebvre P, Silva UL, Prins E (2001) Road paving, fire regime feedbacks, and the future of Amazon forests. For Ecol Manag 154:395–407
45. NRC (National Research Council) (2005) Valuing ecosystem services: toward better environmental decision-making. National Academies, Washington, DC
46. Ostrom E (2005) Understanding institutional diversity. Princeton University Press, Princeton
47. Pagiola S, Arcenas A, Platais G (2005) Can payments for environmental services help reduce poverty? World Dev 33:237–253
48. Pagiola S, Bishop J, Landell-Mills N (2002) Selling forest environmental services. Earthscan, London
49. Parry ML, Canziani OF, Palutikof JP, van de Linden PJ, Hanson CE (eds) (2007) Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK
50. Patz JA, Olson SH (2006) Malaria risk and temperature: influences from global climate change and local land use practices. Proc Natl Acad Sci 103:5635–5636
51. Peterson C, Lubchenco J (1997) Marine ecosystem services. In: Daily G (ed) Nature's services. Island Press, Washington, DC
52. Polasky S, Nelson E, Pennington D, Johnson KA (2010) The impact of land-use change on ecosystem services, biodiversity and returns to landowners: a case study in the state of Minnesota. Environ Resour Econ 48:219–242
53. Postel S, Carpenter S (1997) Freshwater ecosystem services. In: Daily GC (ed) Nature's services. Island Press, Washington, DC, pp 195–214
54. Raudsepp-Hearne C, Peterson GC, Bennett EM (2010) Ecosystem service bundles for analyzing tradeoffs in diverse landscapes. Proc Natl Acad Sci 107:5242–5247
55. Rickenbach MG, Reed AS (2002) Cross-boundary cooperation in a watershed context: the sentiments of private forest landowners. Environ Manage 30:584–594
56. Ricketts T, Regetz J, Steffan-Dewenter I et al (2008) Landscape effects on crop pollination services: are there general patterns? Ecol Lett 11:499–515
57. Rokityanskiy D, Benítez P, Kraxner F et al (2007) Geographically explicit global modeling of land-use change, carbon sequestration, and biomass supply. Technol Forcasting Soc Chang 74:1057–1082
58. Ruhl JB, Kraft SE, Lant CL (2007) The law and policy of ecosystem services. Island Press, Washington, DC
59. Salzman J (2005) Creating markets for ecosystem services: notes from the field. NYU Law Rev 80:870–961
60. Scheffer M, Carpenter SR, Foley J et al (2001) Catastrophic shifts in ecosystems. Nature 413:591–596
61. Schlesinger W (1991) Biogeochemistry: an analysis of global change. Academic, San Diego
62. Tallis H, Goldman R, Uhl M, Brosi B (2009) Integrating conservation and development in the field: implementing ecosystem service projects. Front Ecol Environ 7:12–20

63. Tallis H, Lester SE, Ruckelshaus M, Plummer M, McLeod K, Guerry A, Andelman S, Caldwell MR, Conte M, Copps S, Fox D, Fujita R, Gaines SD, Gelfenbaum G, Gold B, Kareiva P, Kim C, Lee K, Papenfus M, Redman S, Silliman B, Wainger L, White C (2012) New metrics for managing and sustaining the ocean's bounty. Mar Policy 36:303–306
64. Troy A, Wilson M (2006) Mapping ecosystem services: practical challenges and opportunities in linking GIS and value transfer. Ecol Econ 60:435–449
65. United Nations Environment Programme (2006) Marine and coastal ecosystems and human well-being: a synthesis report based on the findings of the millennium ecosystem assessment. UN Environment Programme, Nairobi
66. van Mantgem PJ, Stephenson NL, Byrne JC, Daniels LD, Franklin JF, Fulé PZ, Harmon ME, Larson AJ, Smith JM, Taylor AH, Veblen TT (2009) Widespread increase of tree mortality rates in the Western United States. Science 323:521–524
67. Wallace K (2007) Classification of ecosystem services: problems and solutions. Biol Conserv 139:235–246
68. World Resources Institute (2007) Kenya poverty mapping
69. Wunder S, Engel S, Pagiola S (2008) Taking stock: a comparative analysis of payment for environmental services programs in developed and developing countries. Ecol Econ 65:834–852
70. Zhang P, Shao G, Zhao G et al (2000) China's Forest policy for the 21st century. Science 288:2135–2136

Chapter 7
Ecosystems and Spatial Patterns

Patrick M.A. James and Marie-Josée Fortin

Glossary

Disturbance	A spatial process or event that reverts forest vegetation to early successional stages typically altering forest structure and composition.
Ecotone	A region of interface between two communities, ecosystems, or biogeographic regions.
Legacy	A persisting spatial feature or pattern that was generated by a historical disturbance. Legacies can constrain the spatial dynamics of contemporary disturbances.
Multi-scale analysis	A method of spatial analysis that looks at the relative contributions of different scales of spatial pattern to a single observed spatial pattern.
Pattern	A repeatable and identifiable feature in a spatial context.
Scale	An attribute of a spatial process or data used to represent that process that describes its spatial dimensions. Scale includes elements of grain, extent, and thematic resolution.
Spatial autocorrelation	The degree of correlation of a variable and itself as a function of the spatial distances among sample points.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

P.M.A. James (✉)
Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: patrick.ma.james@umontreal.ca

M.-J. Fortin
Department of Ecology and Evolutionary Biology, University of Toronto,
Toronto, ON, Canada
e-mail: fortinmj@gmail.com

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_7,

Stationarity	A feature of a spatial process in which the mean and variance of a process is consistent across the extent of a study area.
Variography	A geostatistical modeling tool for describing spatial variance and semivariance as a function of spatial distance among pairs of points.

Definition of the Subject

Ecological processes such as forest disturbances act on ecosystems at multiple spatial and temporal scales to generate complex spatial patterns. These patterns in turn influence ecosystem dynamics and have important consequences for ecosystem sustainability. Analysis of ecosystem spatial structure is a first step toward understanding these dynamics and the uncertain interactions among processes. There are many spatial statistics available to describe and test spatial pattern within ecosystems and to infer the character of the processes that generated them. Indeed, improving understanding of the processes that create spatial pattern is a central objective of spatial pattern analysis. In addition to standard tests of spatial autocorrelation and patch structure, methods for multi-scale decomposition of spatial data and identification of stationarity are necessary to determine the key spatial scales at which the processes operate and affect ecosystems and to identify meaningful spatial subunits within larger contexts. Finally, tools for identifying ecosystem boundaries are also important to monitor boundary movement and changes in local ecosystem characteristics through time.

Introduction

Spatial Patterns in Ecosystems

The spatial structure of ecological systems is important to examine and understand as spatial structure mediates the flows of individuals, materials, and information through space and time [1]. These flows bear on the probabilities of occurrence and persistence of floral and faunal populations which determine local and regional biological diversity as well as ecosystem functioning [2]. Interruptions and alterations of such flows within and among ecosystems in terms of rate, quantity, or both as a result of human interventions or natural dynamics such as disturbance can have important consequences for ecosystem sustainability and long-term population persistence. Quantitative characterization of spatial patterns and their rates of change in natural environments is essential to understanding ecological

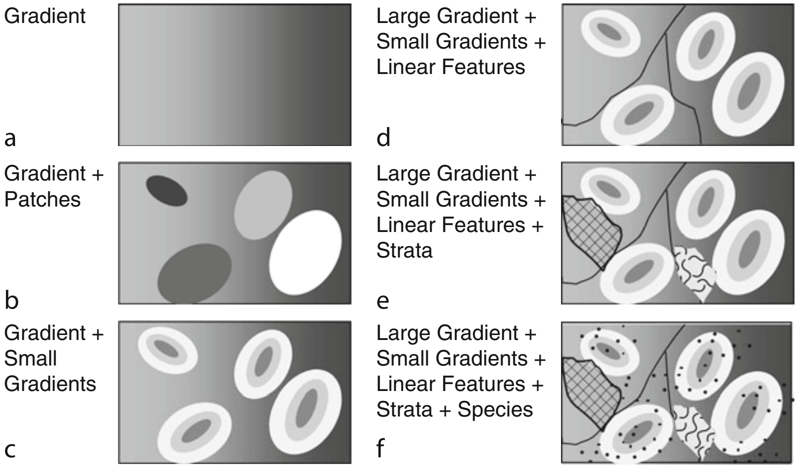


Fig. 7.1 Spatial heterogeneity as a series of additive processes resulting in additive spatial patterns to which individual organisms (here represented points) may respond. Some or all of the different types of spatial heterogeneity may be present in any given landscape

processes and to inform sustainable management techniques that aim to minimize degradation and alteration of ecosystem dynamics [3].

Spatial pattern, or simply spatial structure, refers to a quantifiable attribute of a spatial context. General definitions of the word pattern include a simple definition such as a distinctive or regular "form" or "order," or a feature that is repeated with some degree of "regularity" [4]. Recently, Wagner and Fortin [5] defined the more general term "spatial heterogeneity" as spatially structured variability in a property of interest. Both exogenous environmental (e.g., edaphic variability, elevation, climate) and endogenous ecological (e.g., species interactions, pollen, and seed dispersal) processes generate spatial structure. Each of these two types of spatial processes can produce spatial pattern in multiple forms and scales (Fig. 7.1). The simplest form of spatial pattern is a simple gradient (Fig. 7.1a). Spatial structure can also be present in the form of patches, linear features, and points and can be superimposed on a gradient (Fig. 7.1b–f). When biological spatial structure is mostly responding to environmental conditions such as those depicted in Fig. 7.1, the resulting spatial structure is said to have spatial dependency to the environmental factors. However, when spatial structure emerges as a result of interactions among ecological processes, the pattern is said to be spatially autocorrelated [5].

Spatial patterns and heterogeneity can also be defined using spatial and topological characteristics. These characteristics can include, but are not limited to, a pattern's intensity, autocorrelation, degree of clustering, variability, and scale, which itself includes spatial grain and extent [6]. Importantly, a single pattern summarized using different characteristics can result in different interpretations of the processes behind that pattern [7, 8].

Because spatial pattern analysis is often interested in inferring the processes that created them, it is important to recognize that any single observed pattern represents but one realization of the stochastic process(es) that generated it [7, 9]. By acknowledging that an observed pattern is but a single "snapshot," its temporal dimension is recognized and that under different circumstances, the patterns seen may not be exactly the same. Hence, a main objective of studying spatial pattern is to try to tease apart stochastic processes and the patterns they create from their spatiotemporal conditionalities.

In addition to being driven by processes that are stochastic, patterns emerge as a result of multiple processes that operate at different spatial and temporal scales [10, 11]. These processes can be biotic or abiotic and are usually interconnected through dynamic, and occasionally nonlinear, feedback loops. For example, emergent spatial pattern following forest fires is conditional on the initial distribution of forest fuels as well as fire-weather conditions [12]. Similarly, patterns in forest vegetation composition are often related to patterns in abiotic factors such as moisture, drainage, and soil conditions. Patterns in the genetic composition in animal populations have also been shown to be influenced by the environmental variation (e.g., suitable vs. unsuitable habitat) between sampled populations [13]. These relationships are often nonlinear as the patterns that result from the interactions among pattern-generating processes tend to be different than any one process on its own [14, 15].

Novel spatial patterns created by contemporary anthropogenic processes have uncertain consequences for natural ecosystem dynamics. Anthropogenic processes including deforestation, development, land use change, and climate change do not replace natural processes, but have the capacity to interact with and alter them. As such, a significant question in modern ecology and ecosystem science is that of what are the effects of such novel patterns and processes on natural, or historical, system dynamics [16–18]. Not only do new sources of spatial variability influence natural dynamics through changing the patterns to which natural processes respond, but they also can alter the processes themselves. For example, with regard to forest fire dynamics, this is true where forest composition has been changed due to fire suppression and management (pattern change) and fire frequencies are increased due to increased ignitions near roads or changes in local weather patterns (process change). Similarly, with regard to animal population dynamics, movement and dispersal may be impeded through habitat loss and fragmentation (pattern change) and habitat loss can have an absolute effect on effective population size, rates of dispersal, and genetic variability (process change). Sophisticated spatial statistical analyses are required to begin to disentangle the contributions of different processes to observed spatial patterns to understand how best to manage natural systems to safeguard against further habitat-related losses to biodiversity [19].

Here the causes and consequences of spatial patterns in terrestrial forest ecosystems are reviewed with particular emphasis on patterns of forest vegetation generated through landscape level disturbance processes. Spatial patterns in forest vegetation are both ecologically and economically important in that they are directly relevant to wildlife habitat supply, timber supply, future disturbance dynamics, and represent future challenges to forest and land managers. Uncertainty regarding future disturbance dynamics, in particular fire and insects outbreaks, in the context of global climate change makes investigations into disturbance interactions and potential long-terms consequences for ecosystem spatial structure and functioning particularly relevant.

Forest Ecosystems

In North American forests, disturbance processes generally include landscape level fires, insect outbreaks, forest management (i.e., logging), and fine-scale local disturbances such as windthrow and fungal diseases. The patterns created through the interactions among disturbances can have important economic and ecological consequences. For example, Stadler et al. [20] demonstrated that hemlock wolly adelgid (*Adelges tsugae*) infestations in New England can affect both fast and slow ecosystem dynamics, nutrient cycling dynamics in the short term, and landscape-scale patterns of forest composition in the long term. Similarly, compounded disturbances (e.g., fire and logging) in the eastern boreal forest can result in alternate forest states [21], which can have consequences for biodiversity conservation. Economically, it has been clearly demonstrated that forests under risk of disturbance, either through fire or insect outbreaks, required longer rotation periods to accommodate for the losses [22].

Logging, fire, and insect outbreaks represent disturbance processes that revert forest stands to early seral stages. Succession describes processes of forest recovery, regeneration, and change that vary in response to different disturbances. Although multiple processes generate forest spatial heterogeneity, not all influence it in the same way. Spatial disturbance legacies vary in terms of shape, size, intensity, boundary characteristics, influence on forest succession, and effects on forest age structure [18, 23–25]. The interactions among processes, or more properly, interactions among current disturbance and existing spatial legacies, create and maintain heterogeneous forest landscapes. This cascade of effects and constraints creates mutual dynamic feedbacks among patterns (spatial legacies) and spatial processes (disturbances) [26, 27] with important consequences for ecosystem dynamics.

Different forest disturbances create different forms of spatial structure. Indeed, each disturbance imposes its own unique “spatial signature” on the landscape that also has different temporal characteristics contingent on a disturbance’s interaction with succession (Fig. 7.2). Fires, for example, tend to produce relatively discrete patches that occur over a short time frame and vary in terms of the residual forest structure that is left behind [28]. Logging is somewhat similar to fires in that the patches created are discrete and occur over short time frames and forest managers have control over the scale and amount of residual structure. Insect disturbances, such as outbreaks of spruce budworm (*Choristoneura fumiferana*), forest tent caterpillar (*Malacasoma disstri*), and the mountain pine beetle (*Dendroctonus

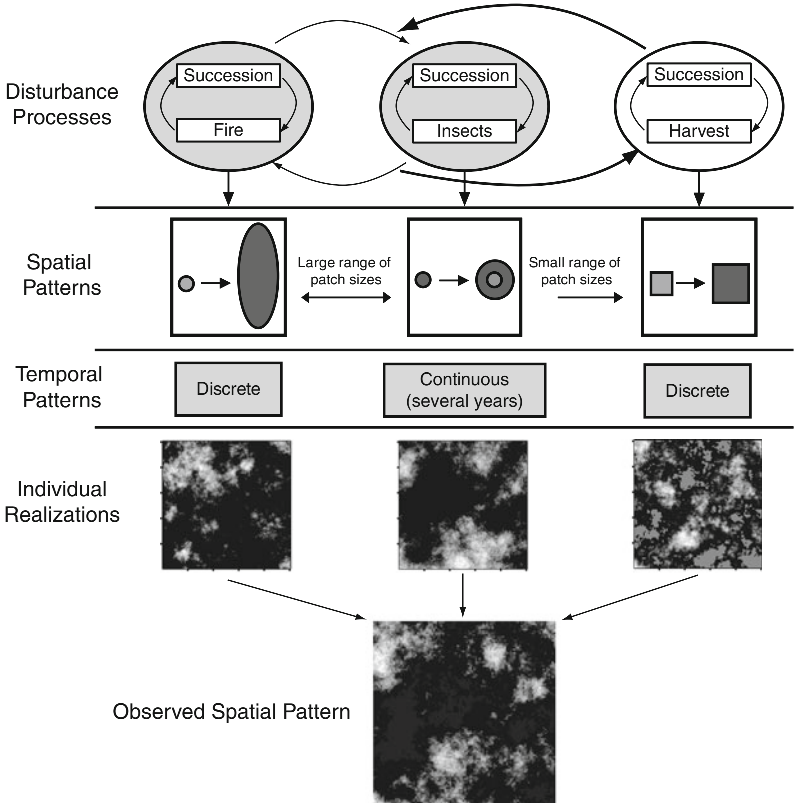


Fig. 7.2 Spatial and temporal scales of forest disturbances. Interactions among disturbances are dependent on the unique successional responses to each disturbance (*row 1*). Historical forest systems were governed by interactions mainly between fire and insects (*arrows*) although presently, logging also interacts with these historical processes. Columns show the unique spatial and temporal attributes of each of the three main boreal forest disturbance agents: fire, insects (i.e., SBW), and harvesting (i.e., logging). These different spatial features result in different realization, or spatial signatures, of each disturbance (*row 4*). Interactions among these different processes produce a single observed spatial realization of spatial structure that contains elements of each of the different processes (*row 5*). Observed patterns contain elements of all three main processes; the objective of spatial analyses is to begin to tease apart the relative contributions of different processes to observed spatial pattern (*row 6*)

ponderosae), are less discrete and tend to produce more complicated spatial structure and continue to affect forest structure at a given location for multiple consecutive years [29].

Each disturbance also has a unique relationship with forest regeneration processes [30] such that forest succession is tightly coupled to the type of disturbance that reinitiates stand development. These relationships determine future forest structure. Historically, fire and insects were the main disturbances in North American forest systems. Adaptations to disturbance such as serotiny in pine species (e.g., jack pine; *Pinus banksiana*) and advanced regeneration in the understory of spruce (e.g., *Picea spp.*) and balsam fir (*Abies balsamea*) stands that maintain spruce budworm host availability over time [31] are evidence of this dynamic feedback between disturbance and succession. The spatial patterns created through forest management and their influences on forest succession in turn influence future forest disturbances dynamics [32]. Spatial pattern analysis is important to better understand the effects of human activities on natural disturbance dynamics.

Sources of Heterogeneity

Understanding the nature and consequences of spatial heterogeneity in ecosystems requires an understanding of the processes that generate this heterogeneity. In this section, different types of spatial heterogeneity and how different types of processes may give rise to complex spatial patterns are described. The consequences and potential challenges involved in indentifying the relative contributions of these different and frequently interacting processes are discussed next [6].

Levels of Organization

Processes that generate spatial heterogeneity can be classified into a hierarchy of spatial processes that operate at different levels of ecological organization: (1) individual, (2) population, (3) community, and (4) landscape/ecosystem. Individual processes include organism dispersal and habitat selection; population processes can include demographic dynamics as well as immigration/emigration; community level processes are highly relevant to natural disturbance dynamics and can include successional changes and rates of species turnover. Examples of landscape/ecosystem level processes include disturbance, climate change, and migration.

Processes within this organizational hierarchy are not necessarily independent and can influence each other among levels. Such interactions can influence emergent patterns due to potential cross-scale interactions and amplifications [33] and can also further complicate efforts to identify clear cause-and-effect relationships. A recent example of such cross-scale amplification in a forest ecosystem can be found within the mountain pine beetle (*Dendroctonus ponderosae*) system of Western Canada, where the recent outbreaks of the lodgepole pine infesting beetle have affected an unprecedented millions of hectares [34]. Here, the local dynamics of population control by host tree defenses were overcome when population numbers increased dramatically due to persistent warmer temperatures in the early 2000s. The interactions between these local- and landscape-level processes are thought to have led to a positive feedback that allowed the outbreak to expand as much as it did [34].

Interactions Among Spatial Processes

Spatial processes within ecosystems interact with each other directly as well as with the spatial legacies of previous, historical processes. In this way, spatial patterns and processes are connected through a dynamic and persistent feedback loop [32]. Spatial legacies can be thought of as a form of spatiotemporal connectivity among disjunct spatial processes or events that are mediated by forest succession and aging. The legacies of historical processes as represented by contemporary patterns can have long-lasting and significant impacts on biodiversity [16, 35] and efforts to sustainably manage forest ecosystems [23, 36].

Spatial legacies can be defined at different scales and describe persisting features within a stand, landscape, or ecosystem. The term "legacy" can refer to fine-scale structural complexity following windthrow [37], landscape level forest age structure [18, 38], disturbance-mediated seed availability [39], and residual forest structure following fire [39, 40]. Because patterns of historical land use can influence contemporary ecosystem composition, configuration, and ecosystem process dynamics long after the actual event [41], a better understanding of spatial legacies and their influence on ecosystem dynamics and landscape change over time is needed and requires novel spatial and temporal methods of investigation and analysis.

From the perspective of sustainability, spatial disturbance legacies, including those created through human activities, represent future ecosystem patterns and future challenges for sustainable management. Gustafson et al. [42] showed that new forest harvest goals are not easy to achieve due to existing conditions when examining shifting forest management rules. Wallin et al. [38] demonstrated that shifts from a dispersed to an aggregated harvest pattern did not immediately result in a change in forest attributes such as patch size and edge density. Instead, new harvest rules had to work around the legacies of previous patterns, and original patterns were enforced. Similarly, Gustafson and Rasmussen [43] found that when varying parameters in a harvest simulation model, the persistent legacies of previous harvest patterns resulted in timber harvest shortfalls. Using a simulation approach, James et al. [18] demonstrated that legacies in forest age structure created through forest management can persist for over 100 years. Ecologically, the consequences of these legacies interacting with new disturbances can result in greater system variability and gradual ecosystem degradation [43–45] or alternative stable states [21, 46].

Multiple Spatial Scales

Inferring the characteristics of spatial processes through analysis of spatial pattern is a central goal of most ecological studies as it is often very difficult to analyze the processes of interest directly. This can be particularly challenging when several types of pattern (Fig. 7.1) and underlying processes are present (Fig. 7.2). The challenge resides in the fact that the single observed pattern is an amalgamation of these multiple processes interacting with existing spatial structure and historical legacies; the functional relationships that connect these contributors to pattern are largely uncertain (Fig. 7.2). When analyzing the spatial structure of sampled data, it is not easy to disentangle the key spatial scales, and therefore processes, that act on the data. However, in the last decade, hierarchical decomposition methods (multiscale ordination [47]; PCNM [48]; wavelets, [49] more detailed below) have been developed to identify the spatial scales at which data are most strongly structured and to decompose the data on the basis of scale-specific variances. Beyond simply describing patterns and the scales at which they are structured, it is also important to have a priori hypotheses about which scales and processes are the most relevant for the questions under study as these methods could reveal many patterns and spatial scales, many of which may not be of relevance [8]. Spatial pattern analysis will be more effective at describing underlying processes when used in an explicit and informed hypothesis testing framework.

Ecological Consequences of Spatial Heterogeneity

The consequences of changes in spatial pattern in forest landscapes are easily confounded by absolute losses in wildlife habitat [19]. That is, although both forest composition and configuration are important, issues related to configuration are only relevant below a critical threshold of forest amount (usually 20–30% of area; [50]). Above the critical threshold, the landscape generally remains “connected” and organisms or disturbances can spread in the landscape [51]. Below such a critical threshold, species’ response to the amount of habitat area is nonlinear as most species do not have enough habitat to meet their needs. Surrounding habitat quality (composition) and configuration become more important for local population persistence in this case. Moreover, fragmented landscapes with various landcover types can impede species abilities to move from habitat patches to another [52]. For example, nesting birds do not cross forest gaps larger than 25 m [53]. Impediments to movement across landscapes can influence population dynamics [54] as well as genetic heterogeneity [55], both of which affect the probability of population persistence.

Spatial Analyses

There are three main approaches to investigating the different aspects and consequences of spatial heterogeneity. Spatial statistics, landscape metrics, and statistical modeling, all approach the question of identifying spatial pattern in ecosystems in a slightly different way [5]. Owing to the varied history of approaches to studying spatial patterns including methods and concepts drawn from a diverse set of disciplines such as geography, geostatistics, and ecology, only the key concepts related to the most commonly used methods that quantify spatial structure within and among ecosystems are presented. Before describing the different types of analysis, some fundamental issues related to the spatial analysis of data are presented.

Assumption of Stationarity

To infer spatial pattern from samples spatial statistics require that the area under study is governed by the same underlying process (i.e., the assumption of stationarity; [7]). As it is often impossible to be sure that the underlying process is stationary, one needs either to assume it or to determine whether or not the observed data are stationary (i.e., their statistical properties such as mean, variance, isotropy do not vary with spatial distance). Nonstationary processes may arise when more than one process is present and that these multiple processes may be acting at different spatial or temporal scales. Yet, in most forest ecosystems, processes interact with one another, resulting in unique types and scales of spatial pattern which violate the assumption of stationarity. In such circumstances, it is required to first identify stationary subregions within such a larger spatial context. A few spatial analysis methods do not require the stationarity such as lacunarity analysis, local quadrat variance methods, and wavelets [7]. It is worth noting that these types of analyses although different and originating from different developmental histories are quite similar to one another mathematically [56].

Data Type

Spatial pattern within ecosystems can be represented using categorical or continuous data depending on the nature of the variable under investigation. Each type of data requires different methods of analysis (Fig. 7.3; [7]). Categorical data can be described by the amount and configuration of the different discrete types on the landscape. Examples of categorical spatial data include forest type and age, or classified habitat patches. Both amount and configuration can be described in numerous ways using landscape pattern metrics [9, 57, 58]. Continuous data requires more subtlety in describing patterns and can include variables such as soil moisture, forest basal area, or remotely sensed reflectance indices (e.g., NDVI). Composition of continuous variables can be described using the density distribution of the variable and configuration is usually described by a spatial covariance function that captures the strength, directionality, and scale of autocorrelation of the variable [5, 7, 59].

In addition to the categorical/quantitative dichotomy of data types, patterns can be described using different geometric topologies of spatial features or units: the

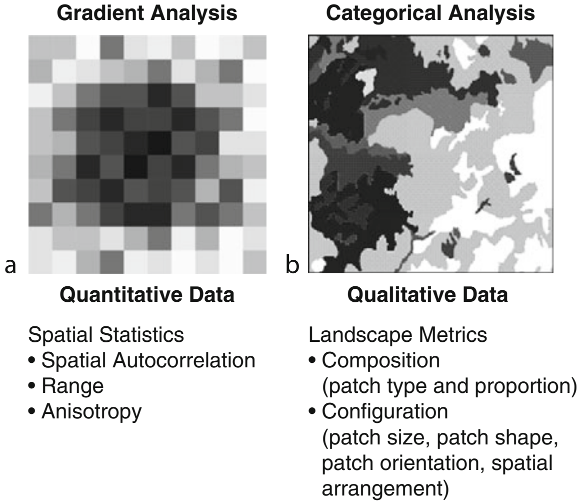


Fig. 7.3 Spatial analysis can be undertaken on different types of spatial data. (**a**) Raster-based quantitative spatial data (e.g., forest height, basal area, NDVI) that can be analyzed using spatial statistics to determine the intensity, spatial range, and directionality (anisotropy) of the spatial pattern. (**b**) Categorical and qualitative forest data (e.g., species, stand age) require a different analytical approach that typically includes landscape pattern metrics

vector format (points, lines, polygons) and the raster (i.e., pixel) format [60]. Representation of spatial structure in one form of data does not preclude use of another form. For example, annual polygons of insect defoliation data can be converted into raster form and analyses can be undertaken using time series of raster values at a specific location [61]. Furthermore, binary rasters (presence/absence) can be converted to continuous rasters by increasing the cell size and counting the number of "presence" pixels surrounding a focal pixel using 4-neighbour, 8-neighbour, 16-neighbour rules and assigning this value to the to the new, larger pixel.

Raster data can be used to represent any continuous variable. In contrast to point data, in raster data types, the information fully covers the extent of the study area. There is also a unique grain (or cell size) to each raster "pixel" that determines the subarea of continuous space that is discretized by the raster cell. The selection of raster grain can have important consequences to the results of spatial analyses [6]. Raster data can be used to represent any number of spatial variables relevant to disturbance ecology including, but not limited to, tree species [62], stand age [18, 63], basal area [64], insect damage [61], and number of fire occurrences [65]. The continuous coverage of raster data makes it amenable to many different analytical techniques such local quadrat variance, lacunarity, and wavelets [56].

Spatial Analyses within an Ecosystem

Ecological variables that are geographically distributed in space and time tend to be more similar when compared close together [66]. Autocorrelation is a feature of most data and can be quantified by the degree of self-similarity or dissimilarity in a variable between pairs of locations at a given distance apart (i.e., spatial lag determined in terms of equidistant classes). Note that spatial statistics on their own cannot differentiate between spatial dependence to environmental factors and spatial autocorrelation due to ecological processes; only prior knowledge and multiple testing can differentiate between these two sources of spatial structure.

Spatial Description of the Pattern

The objective of many spatial statistics is often to characterize to what degree spatial data are autocorrelated, if they are oriented in a particular direction (anisotropy), and at what scale. As these spatial statistics have been thoroughly reviewed elsewhere [7, 58], we focus on three topics (1) methods of spatial pattern analysis devoted to identifying structure in point data, (2) methods of spatial analysis that are devoted to identifying structure and pattern in two-dimensional raster (pixel) or polygon data, and (3) methods of spatial analysis explicitly concerned with identifying the scale, or scales of structure that are present in either point or two-dimensional data. Both the data types discussed can be examined in uni-, bi-, and multivariate contexts and can include either categorically or continuously measured variables.

Point Pattern Analysis

Spatial point processes describe phenomena that produce events represented as points in space [67, 68]. The objective of point pattern analysis is to determine whether the distribution of events (points in space) is more or less spatially aggregated than is expected by chance and tests the null hypothesis of complete spatial randomness. The use of complete spatial randomness assumes that the underlying process is the same over the study area (i.e., stationarity). When it is not the case, the study area is said to “inhomogeneous” such that significance cannot be achieved using a single process such as complete spatial randomness. Modified statistics and corrections have been developed to account for inhomogeneity within the study area [69, 70]. Point pattern analysis also assumes complete census of all point occurrences in the study area [7]. Fortin et al. [71] showed that the significance of spatial aggregation estimations is biased only when a subset of sample points is used rather than the entire set of points in the study area.

The most commonly used method of point pattern analysis is Ripley's *K* statistic [72]. New statistics have also been developed to compute local estimates [68]. Ripley's *K* statistic, and derived statistics, can be applied in one-, two-, or three-dimensional space to compare the degree of aggregation of points [68]. In cases where two (or more) point processes are operating, it may be of interest to assess whether one process influences the other and whether points of different types tend to cluster together. Examples of ecologically relevant spatial point processes include fire occurrence, plant occurrences [67], or the distribution of animal nesting or denning sites [72]. The bivariate, or cross-*K*, Ripley's *K* test assesses whether the co-occurrence of two types of points is clustered together more or less than is expected by chance [58]. Using this technique, Lynch and Moorcroft [73] examined co-occurrence of fire and insect outbreaks and found that contrary to expectation, insect-caused forest mortality does not increase the risk of forest fire.

Spatial Autocorrelation

Often, a researcher is interested in determining the scale and strength of spatial autocorrelation of a variable as well as whether there is a directional trend in the data (i.e., anisotropy). This can be achieved using spatial autocorrelation coefficients such as Moran's *I* which computes the product of the deviations of the values of the variable to its average according to various distance intervals (lags, classes) standardized by the variance at that spatial lag [7]. Moran's *I* behaves like a Pearson's correlation coefficient such that the null hypothesis is the absence of spatial autocorrelation, positive autocorrelation (mostly a short distance) indicates that values have comparable values, while negative values indicate that the values are very dissimilar. Moran's *I* assumes that the underlying process is the same over the entire study area (i.e., stationarity). Hence, the spatial autocorrelation coefficients computed at various distance classes are average values. Spatial autocorrelation in this sense can be referred to as a global spatial statistic that describes an attribute of the data over the entire study area [7]. Significance of each coefficient can be computed based on an asymptotic *t* or randomization procedure. In either case, stationarity is required. Moran's *I* is very sensitive to skewed data as the mean will be biased and in consequence all the deviations values based on it will also be biased. It is therefore recommended to check the distribution of the data before computing spatial autocorrelation and if needed transform the data to obtain a symmetric distribution.

Measures of spatial autocorrelation are also sensitive to sample size. When autocorrelation is estimated using too few locations, (e.g., <30 positions), spatial patterns may not be detected, even though present. Similarly, depending on the spacing among sampling locations, there will be different numbers of paired comparisons at each lag distance. Typically, there are few pairs at short distances due to the edge effects of the edge of the study area (no locations outside to compare too), most of the pairs at intermediate distances and very fewer at large distances (because of the overall size of the study area). To mitigate these unequal numbers of pair per distance lag, it is recommended to focus on lags distances equal to the length of the first half (up to two thirds) of the smallest edge of the study area as most spatial autocorrelation occurs at short distances and that the probability of detecting it is also highest in the first spatial lag [7].

A plot of spatial autocorrelation coefficients against distance lags is called a spatial correlogram. If the lags are based on distance only, the correlogram is said to be an omnidirectional correlogram. When the data contain directionality, the omnidirectional correlogram cannot reveal it and may, in fact, "mask" it. To detect the presence of anisotropic spatial pattern (i.e., not having the same sill and range according to direction), the samples need to be divided by distance class as well as direction angle range (usually 0°, 45°, 90°, 135°) to produce a set of directional correlograms.

In areas where several processes influence ecological data, Moran's I that assumes stationarity cannot be used. Instead, local indicator of spatial aggregation statistics, LISA (e.g., local Moran, local Getis), can be used as they are computed at each sampling location and allow the identification of subareas that have similar high ("hot spots") or low ("cold spots") values [7].

Geostatistics

Spatial structure can be determined in terms of spatial autocorrelation as presented above or as spatial variance according to distance as computed using variograms which are part of the family of spatial statistical methods known as geostatistics [7, 58]. Variograms represent a global method of scale-specific analysis that has been used extensively in ecology to analyze spatial patterns [58]. Variograms model the relationship between lag distance and semivariance and can be calculated using continuous raster or point data. Semivariance is calculated as the sum of the squared differences between pairs of locations separated by a given lag distance divided by twice the number of pairs of locations at that particular lag distance [7]. From the observed, or empirical, variogram, three parameters can be estimated to fit a theoretical variogram: (1) range, or scale at which distance does not affect the estimate of variance, (2) sill, or the variance of the data, and (3) nugget effect, which represents the variability in the data that is not accounted for by spatial structure [58]. Theoretical variogram models can identify whether there is a directional trend in the data, that is, anisotropy as the spatial autocorrelation values. In addition to describing the attributes of spatial structure, geostatistical models can be used to "krige" (i.e., spatially interpolate) data [58] and to simulate spatial patterns using a chosen variogram model, process model, and parameter estimates [74].

Spatial Scale and Scaling

In addition to understanding the type (e.g., trend, patch) and strength (e.g., degree and distance) of a spatial pattern, it is useful to identify the spatial scale at which such patterns are present. Because patterns are the result of multiple processes that each have their own unique scales of spatial structure [75], disentangling the relative contributions of these processes and assigning relative importance to them and the scales at which they operate are of fundamental importance to ecology and improving the understanding of complex systems [11, 16]. Furthermore, the identification of the relative contributions of different processes and scales to observed patterns is necessary for understanding cross-scale interactions [33] which is necessary to make reliable predictions of system dynamics, ostensibly the objective of any spatial analysis [76].

As stated above, spatial pattern describes a "quantifiable attribute of a spatial context." Scale-specific analysis identifies the spatial scale at which that attribute is structured establishes its specific context. With regard to forest disturbance dynamics, for example, individual forest stands may seem unstable through the processes of destruction and renewal through disturbances such as fire, but the larger forest landscape (i.e., collection of stands) is in fact stable with respect to the proportion and relative configuration of the different stand types. This is what is meant when disturbance-mediated forest systems are described as a shifting mosaic [77]. Different conclusions would be drawn about forest stability and resilience (sensu [78]) depending on the spatial or temporal scale of investigation.

Scale generally describes the spatial extent, grain, and thematic resolution of a set of data [6]. However, scale can also be used to refer to a level within an organizational hierarchy to which such data pertain, such as a population, or community, or ecosystem [79]. It is important to note that scale in this latter sense is not directly equivalent to the former; scaling up, that is, increasing from local to a broader extent, or aggregating data from a fine to a coarse scale may move the analysis into another level or an organizational hierarchy, but not necessarily [80]. Although scale is best thought of as a feature of the phenomena of interest, it can also be a feature of sampling scheme imposed by a researcher, or the methods of analysis applied [6]. All of these features can influence a researcher's ability to identify scales of structure in spatial data and to make meaningful inferences regarding the underlying processes. It is therefore important to distinguish structure that is emergent from the data from those related to sampling or analytic scales (i.e., arbitrary scales; [10]), as such a priori scales may have little to do with the actual scale of structure in the ecological phenomena of interest [76].

The ability to identify meaningful scales of spatial structure depends on the methods used and the type of data being analyzed [77, 81]. Methods differ in their ability to identify local vs. global scales of pattern. Global methods of scale-specific analysis summarize spatial pattern at a single scale and generally assume that the underlying processes are stationary. Examples of global methods of analysis include variography [58], spectral (i.e., Fourier) analysis [82], and global measures

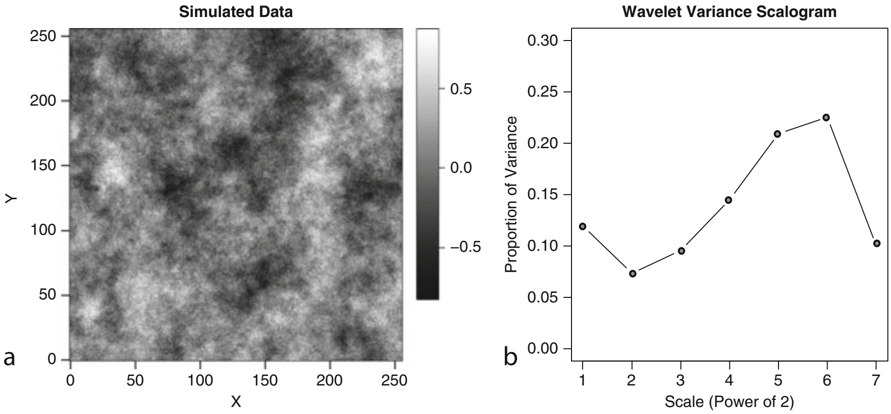


Fig. 7.4 Example of a hierarchical multi-scale decomposition of two-dimensional, quantitative data using wavelets. (**a**) Simulated spatial data. (Data were simulated using an exponential variogram model with the following parameters: Sill = 1; Range = 40; Nugget = 0.1). (**b**) Scalogram that summarizes the proportion of total variance in the original data associated with each scale of the decomposition. Wavelet decomposition was accomplished using a maximal-overlap discrete wavelet transform (MODWT; Percival and Walden [109])

of spatial autocorrelation such as Moran's I and Geary's c [7]. Multi-scale methods of analysis identify both global and local scales of structure, can assign relative importance to difference scales, and do not assume stationarity. Instead, such methods can be used to identify boundaries and scale-specific stationary subregions within a larger spatial context [7, 74]. Multi-scale methods of analysis include lacunarity analysis [83], wavelets [74, 81, 84], distance-based Eigenvector methods (e.g., PCNM; [85]), and local spatial statistics [86].

Wavelet analysis is a particularly powerful method of local spatial analysis that can be used to decompose continuous data into its scale-specific components [87, 88]. A proportion of the total variance in the data set is associated with each level of the decomposition through use of the wavelet variance [84] and the relative contributions of different scales to overall structure can be assessed and visualized as a scalogram (Fig. 7.4b). Each of these scales of pattern can then be isolated using a multi-resolution decomposition (Fig. 7.5; [88]). Under conditions where observed spatial pattern is assumed to be the result of multiple interacting processes, such data decomposition provides an opportunity to assess the relationship among processes and individual scales of spatial structure present in the data. In combination with the scalogram (Fig. 7.4b), the relative importance of these different scales can also be determined and further analyses can be restricted to only those spatial layers that correspond to the scales of interest. Isolated scales of spatial pattern can then be examined independently or used as scale-specific predictors in further statistical analyses [89].

Whereas Fourier analysis assumes that observed patterns can be described as a sum of sine waves of different frequencies, wavelet analysis identifies global and local structure at different scales using a local wavelet template that can take on

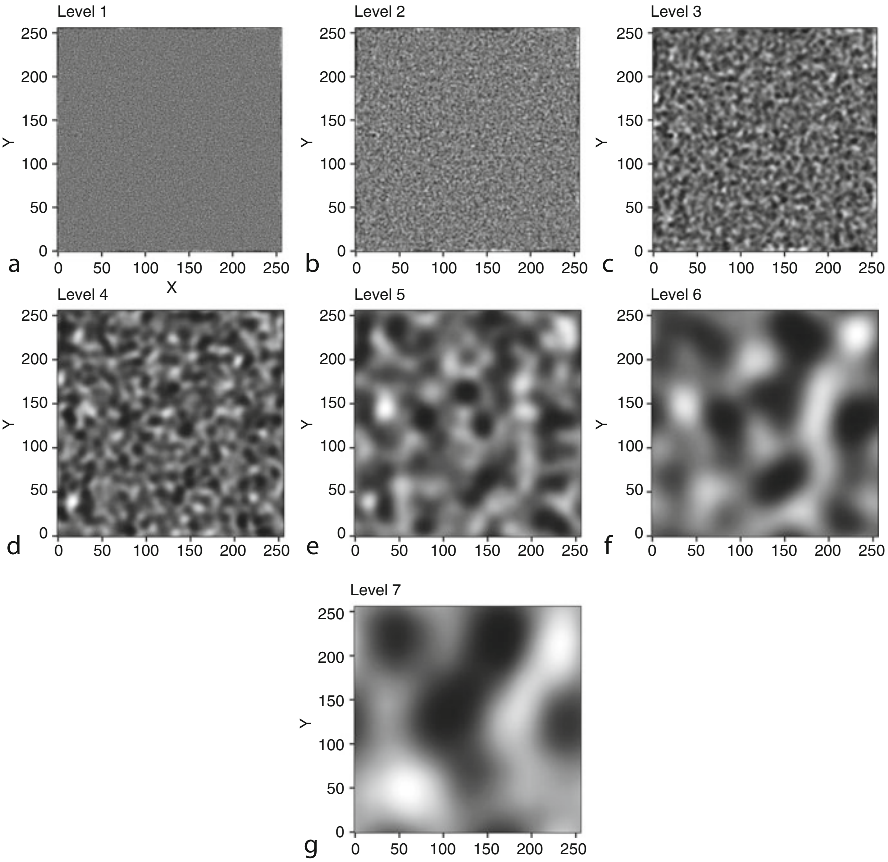


Fig. 7.5 Back-transformed images from the decomposition of the data in Fig. 7.4a. Each panel shows the spatial structure at an individual isolated scale

a wide variety of shapes and forms [90]. The majority of wavelet applications have been in the analysis of temporal signals to identify periodicity in things such as climatic variability [91] and epidemiological time series [92, 93]. However, spatial applications of wavelet analysis in ecology continue to be developed and have been used to investigate one-dimensional forest canopy gap structure [87], vegetation reflectance [89], two-dimensional structure in grassland productivity [81], tree crown identification [94], and the significance of spatial structure in forest basal area [74].

When data are not sampled in a continuous way (i.e., they are irregularly spaced), multi-scale decompositions can be performed using spatial eigenfunction analyses such as principal coordinate analysis of neighbor matrices (PCNM) and Moran's eigenvector maps (MEM) [48, 95, 96]. These methods model spatial structure in a multivariate framework using distance matrices where the spatial coordinates of the sampled sites are converted into a set of synthetic spatial variables that represent spatial structure at different spatial scales. These synthetic variables can then be used in combination with constrained ordination methods (e.g., canonical correspondence analysis and redundancy analysis; [95]) to identify the spatial scales at which data are dominantly structured. Jombert et al. [97] also proposed a multi-scale pattern analysis (MSPA) to determine which of these many scales are the most relevant to use as spatial predictors in subsequent analyses such as partial ordination or multiple regression [85].

Spatial Analyses Among Ecosystems

The ecotonal interfaces between ecosystems are important to delineate as they are the locations where the exchange of nutrients and species turnover occurs [7]. It is also important to determine not only the boundary location between ecosystems, that is, where one system begins and another ends, but also its width (i.e., sharp/line or gradual/zone) [98]. There are two different types of methods that can be used to determine the interface between ecosystems: by creating spatial clusters or by detecting boundaries [99]. In either case, the sampling design used to collect the data is crucial: The determination of a boundary (i.e., area of high rate of changes in the values of variables, hence a heterogeneous area) is relative to the two adjacent spatially homogeneous ecosystems. Therefore, the sampled data should cover enough of both ecosystems such that their interface can be detected.

Spatial Clustering

To delimit boundaries between ecosystems, spatially homogeneous clusters can be determined based on the degree of similarity of sample attributes and their spatial adjacencies [66, 99, 100]. The degree of similarity can be based on commonly used clustering algorithms (agglomerative, *k*-means, fuzzy logic, etc.) and adjacency can be based on network connectivity algorithms (e.g., nearest neighbors, minimum spanning tree, Gabriel network, Delaunay network; [7]). Spatial clustering provides the membership of each location to a spatial cluster and therefore, as a by-product, identifies boundaries among clusters. However, clustering procedures do not provide any information about the location and width of the identified boundaries between clusters. When information about location and width are needed, other methods should be used such as boundary detection methods ([99]; but see [101]).

Spatial Boundary Analyses

Ecological boundaries can be defined as areas of high rates of change or large absolute differences between adjacent locations [102]. Boundary detection methods [98] include edge detection algorithms (Laplacian, Canny, Sobel, Monomier, etc.), wavelet analysis [74, 81], and wombling (lattice, triangulation, categorical). The two former families of methods require the data to be in a contiguous fashion (grid) without any missing values while the latter one can either be used with contiguous or irregularly sampled data. All methods compute the magnitude of rate of change between adjacent locations over the entire extent of the study area. To determine which rates of changes are significantly higher than the others, different statistics and significance tests having been developed. Wombling typically uses arbitrary percentile thresholds of boundary elements to identify significant boundaries [7]. James et al. [74] proposed a series of restricted randomization procedures to test the significance of wavelet boundaries using variogram-based spatial null models. Oden et al. [103] developed boundary statistics to test the cohesiveness properties of the boundaries. Once cohesive boundaries have been detected and tested, subsequent hypothesis testing can be performed by comparing the spatial overlap and movement of boundaries using spatial overlap statistics [7, 104, 105] or polygon change analysis [106].

Future Directions

The detection and characterization of spatial structure is necessary for developing the understanding of ecosystem function and for ensuring sustainable management and use of landscape resources. Indeed, the current scale and pace of anthropogenic influence on the natural environment is without precedent and the ways in which novel human-created spatial patterns interact with and influence natural processes are uncertain. Even more uncertain are the relationships among different scales of spatial and temporal pattern and what such cross-scale interactions may mean to ecosystem dynamics [107]. Future directions in the analysis of spatial pattern will require approaches and methods that can begin to tease apart the separate scales, both spatial and temporal, of processes that contribute to spatial patterns both within and among ecosystems. These methods should include multi-scale approaches such as the wavelet-based methods described above, tools to assess statistically significant changes over time, and methods that can identify local spatially significant subregions within larger spatial contexts [74]. Meaningful inference in these regards will only be possible if the spatial pattern analysis is applied within a hypothesis testing framework, where competing notions of how individual processes percolate through the landscape to produce pattern can be tested statistically [8]. Finally, increasing availability of remotely sensed data (e.g., NDVI, LiDAR, Quickbird, LANDSAT, MODIS) will allow people to detect spatial patterns at finer spatial and temporal scales over much larger spatial extents than has been previously possible. These huge amounts of data will also require focused, hypothesis-driven questions, the use of data-mining tools [108], and the further development of spatiotemporal statistics.

Acknowledgments This work was funded by a Killam Postdoctoral fellowship at the University of Alberta to PMAJ and an NSERC Discovery grant to MJF.

Bibliography

Primary Literature

1. Lindenmayer DB, Fischer J (2006) Habitat fragmentation and landscape change. Island Press, Washington, DC
2. Loreau M, Naeem S, Inchausti P et al (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. Science 294:804–808
3. Lindenmayer DB, Franklin JF (2002) Conserving forest biodiversity. A comprehensive multiscaled approach. Island Press, London
4. Epperson BK (2003) Geographical genetics. Princeton University Press, Princeton
5. Wagner HH, Fortin M-J (2005) Spatial analysis of landscapes: concepts and statistics. Ecology 86:1975–1987
6. Dungan JL, Perry JN, Dale MRT et al (2002) A balanced view of scale in spatial statistical analysis. Ecography 25:626–640
7. Fortin M-J, Dale MRT (2005) Spatial analysis. Cambridge University Press, Cambridge
8. McIntire EJB, Fajardo A (2009) Beyond description: the active and effective way to infer processes from spatial patterns. Ecology 90:46–56
9. Fortin M-J, Boots B, Csillag F, Remmel TK (2003) On the role of spatial stochastic models in understanding landscape indices in ecology. Oikos 102:203–212
10. Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3:385–397
11. Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73:1943–1967
12. Flannigan MD, Stocks BJ, Turetsky MR, Wotton BM (2008) Impact of climate change on fire activity and fire management in the circumboreal forest. Global Change Biol 14:1–12
13. McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. Proc Natl Acad Sci USA 104:19885–19890
14. Veblen TT, Hadley KS, Nel EM et al (1994) Disturbance regime and disturbance interactions in a Rocky Mountain subalpine forest. J Ecol 82:125–135
15. Radeloff VC, Mladenoff DJ, Boyce MS (2000) Effects of interacting disturbances on landscape patterns: Budworm defoliation and salvage logging. Ecol Appl 10:233–247
16. Levin SA (2000) Multiple scales and the maintenance of biodiversity. Ecosystems 3:498–506
17. Cooke BJ, Nealis VG, Regniere J (2007) Insect defoliators as periodic disturbances in northern forest ecosystems. In: Johnson EA, Miyanishi K (eds) Plant disturbance ecology: the process and the response. Elsevier, Amsterdam, pp 1–10
18. James PMA, Fortin M-J, Fall A, Kneeshaw D, Messier C (2007) The effects of spatial legacies following shifting management practices and fire on boreal forest age structure. Ecosystems 10:1261–1277
19. Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Ann Rev Ecol Evol Syst 34:487–515
20. Stadler B, Müller T, Orwig D, Cobb R (2005) Hemlock woolly adelgid in New England forests: canopy impacts transforming ecosystem processes and landscapes. Ecosystems 8:233–247
21. Jasinski JPP, Payette S (2005) The creation of alternative stable states in the southern boreal forest, Quebec, Canada. Ecol Monogr 75:561–583
22. MacLean DA, Erdle TA, MacKinnon WE et al (2001) The spruce budworm decision support system: forest protection planning to sustain long-term wood supply. Can J Forest Res 31: 1742–1757
23. Franklin JF, Forman RTT (1987) Creating landscape patterns by forest cutting: Ecological consequences and principles. Landscape Ecol 1:5–18
24. Fall A, Fortin M-J, Kneeshaw DD et al (2004) Consequences of various landscape-scale ecosystem management strategies and fire cycles on age-class structure and harvest in boreal forests. Can J Forest Res 34:310–322

25. Greene DF, Gauthier S, Noe J, Rousseau M, Bergeron Y (2006) A field experiment to determine the effect of post-fire salvage on seedbeds and tree regeneration. Front Ecol Environ 4:69–74
26. Watt AS (1947) Pattern and process in the plant community. J Ecol 25:1–22
27. Turner MG (1989) Landscape ecology: the effect of pattern on process. Ann Rev Ecol Evol Syst 20:171–197
28. Bergeron Y, Leduc A, Harvey BD, Gauthier S (2002) Natural fire regime: a guide for sustainable management of the Canadian boreal forest. Silva Fenn 36:81–95
29. Boulanger Y, Arseneault D (2004) Spruce budworm outbreaks in eastern Quebec over the last 450 years. Can J Forest Res 34:1035–1043
30. Schroeder D, Perera AH (2002) A comparison of large-scale spatial vegetation patterns following clearcuts and fires in Ontario's boreal forests. Forest Ecol Manage 159:217–230
31. Baskerville GL (1975) Spruce budworm: super silviculturist. Forestry Chron 51:138–140
32. Turner MG (1993) A revised concept of landscape equilibrium: disturbance and stability on scaled landscapes. Landscape Ecol 8:213–227
33. Peters DPC, Pielke RA, Bestelmeyer BT et al (2004) Cross-scale interactions, nonlinearities, and forecasting catastrophic events. Proc Natl Acad Sci 101:15130–15135
34. Raffa KF, Aukema BH, Bentz BJ et al (2008) Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. Bioscience 58:501–517
35. Peterson GD (2002) Contagious disturbance, ecological memory, and the emergence of landscape pattern. Ecosystems 5:329–338
36. Kohm KA, Franklin JF (1997) Creating a forestry for the 21st century. Island Press, Washinton, DC
37. Foster DR, Knight DH, Franklin JF (1998) Landscape patterns and legacies resulting from large, infrequent forest disturbances. Ecosystems 1:497–510
38. Wallin DO, Swanson FJ, Marks B (1994) Landscape pattern response to changes in pattern generation rules: land-use legacies in forestry. Ecol Appl 4:569–580
39. Tang SM, Franklin JF, Montgomery DR (1997) Forest harvest patterns and landscape disturbance processes. Landscape Ecol 12:349–363
40. McIntire EJB (2004) Understanding natural disturbance boundary formation using spatial data and path analysis. Ecology 85:1993–1943
41. Dupouey JL, Dambrine E, Lafitte JD, Moares C (2002) Irreversible impact of past land use on forest soils and biodiversity. Ecology 83:2978–2984
42. Gustafson EJ (1996) Expanding the scale of forest management: allocating timber harvests in time and space. Forest Ecol Manage 87:27–39
43. Gustafson EJ, Rasmussen LV (2002) Assessing the spatial implications of interactions among strategic forest management options using a Windows-based harvest simulator. Comput Electron Agric 33:179–196
44. Paine RT, Tegner MJ, Johnson EA (1998) Compounded perturbations yield ecological surprises. Ecosystems 1:535–545
45. Folke C, Carpenter S, Walker B et al (2004) Regime shifts, resilience, and biodiversity in ecosystem management. Ann Rev Ecol Evol Syst 35:557–581
46. Schröder A, Persson L, De Roos AM (2005) Direct experimental evidence for alternative stable states: a review. Oikos 110:3–19
47. Wagner H (2004) Direct multi-scale ordination with canonical correspondence analysis. Ecology 85:342–351
48. Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826–1832
49. Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Ecology 86:2497–2504
50. Andrén H (1994) Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat – a review. Oikos 71:355–366
51. Gardner RH, Milne BT, Turnei MG, O'Neill RV (1987) Neutral models for the analysis of broad-scale landscape pattern. Landscape Ecol 1:19–28

52. O'Brien D, Manseau M, Fall A, Fortin M-J (2006) Testing the importance of spatial configuration of winter habitat for woodland caribou: An application of graph theory. Biol Conserv 130:70–83
53. Bélisle M, Desrochers A, Fortin M-J (2001) Influence of forest cover on the movements of forest birds: a homing experiment. Ecology 82:1893–1904
54. Fahrig L, Merriam G (1985) Habitat patch connectivity and population survival. Ecology 66:1762–1768
55. Bohonak AJ (1999) Dispersal, gene flow, and population structure. Q Rev Biol 74:21–45
56. Dale MRT, Dixon P, Fortin M-J et al (2002) Conceptual and mathematical relationships among methods for spatial analysis. Ecography 25:558–577
57. Gustafason EJ (1998) Quantifying landscape spatial pattern: what is state of the art? Ecosystems 1:143–156
58. Cushman SA, McGarigal K (2008) Landscape metrics, scales of resolution. In: von Gadow K, Pukkala T (eds) Designing green landscapes. Springer, New York, pp 33–52
59. Cressie NA (1993) Statistics for spatial data. Wiley, New York
60. Burrough PA, McDonnell RA (1998) Principles of geographical information systems. Oxford University Press, Oxford
61. Candau J-N, Fleming R (2005) Landscape-scale spatial distribution of spruce budworm defoliation in relation to bioclimatic conditions. Can J Forest Res 35:2218–2232
62. Wolter PT, White MA (2002) Recent forest cover type transitions and landscape structural changes in northeast Minnesota, USA. Landscape Ecol 17:133–155
63. Didion M, Fortin M-J, Fall A (2007) Forest age structure as indicator of boreal forest sustainability under alternative management and fire regimes: a landscape level sensitivity analysis. Ecol Model 200:45–58
64. Townsend PA, Foster JR, Chastain RA, Currie WS (2003) Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian mountains using Hyperion and AVIRIS. IEEE Trans Geosci Remote Sens 41:1347–1354
65. Wotton BM, Martell DL (2005) A lightning fire occurrence model for Ontario. Can J Forest Res 35:1389–1401
66. Legendre P, Fortin M-J (1989) Spatial pattern and ecological analysis. Vegetatio 80:107–138
67. Dale MRT (1999) Spatial pattern analysis in plant ecology. Cambridge University Press, Cambridge
68. Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns. Wiley-Interscience, Chichester
69. Baddeley AJ, Moller J, Waagepetersen R (2000) Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica 54:329–350
70. Wiegand T, Moloney KA (2004) Rings, circles and null-models for point pattern analysis in ecology. Oikos 104:209–229
71. Fortin M-J, Dale MRT, Bertazzon S (2010) Spatial analysis of wildlife distribution and disease spread. In: Huettmann F, Cushman SA (eds) Spatial complexity, informatics, and wildlife conservation. Springer, New York, pp 255–273
72. Ripley BD (1976) The second order analysis of stationary point processes. J Appl Probab 13:255–266
73. Lynch HJ, Moorcroft PR (2008) A spatiotemporal Ripley's K-function to analyze interactions between spruce budworm and fire in British Columbia, Canada. Can J Forest Res 38:3112–3119
74. James PMA, Fleming RA, Fortin M-J (2010) Identifying significant scale-specific spatial boundaries using wavelets and null models: spruce budworm defoliation in Ontario, Canada as a case study. Landscape Ecol 25:873–887
75. Holling CS (1992) Cross-scale morphology, geometry, and dynamics of ecosystems. Ecol Monogr 62:447–502
76. Gardner RH (1998) Pattern, Process, and the Analysis of Spatial Scales. In: Peterson DL, Parker VT (eds) Ecological scale. Columbia University Press, New York, pp 17–34

77. Saunders SC, Chen J, Drummer TD et al (2002) The patch mosaic and ecological decomposition across spatial scales in a managed landscape of northern Wisconsin, USA. Basic Appl Ecol 3:49–64
78. Holling CS (1973) Resilience and stability of ecological systems. Ann Rev Ecol Evol Syst 4:1–23
79. Allen TFH, Hoeskstra TW (1992) Toward a unified ecology. Columbia University Press, New York
80. O'Neill RV, Krummel JR, Gardner RH et al (1988) Indices of landscape pattern. Landscape Ecol 1:153–162
81. Csillag F, Kabos S (2002) Wavelets, boundaries, and the spatial analysis of landscape pattern. Ecoscience 9:177–190
82. Keitt TH (2000) Spectral representation of neutral landscapes. Landscape Ecol 15:479–494
83. Saunders SC, Brosofske KD, Chen J, Drummer TD, Gustafson EJ (2005) Identifying scales of pattern in ecological data: a comparison of lacunarity, spectral and wavelet analyses. Ecol Complex 2:87–105
84. Dale MRT, Mah M (1998) The use of wavelets for spatial pattern analysis in ecology. J Veg Sci 9:805–814
85. Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Model 153:51–68
86. Anselin L (1995) Local indicators of spatial association - LISA. Geogr Anal 27:93–115
87. Bradshaw RHW, Spies T (1992) Characterizing canopy gap structure in forests using wavelet analysis. J Ecol 80:205–215
88. Mallat S (1999) A wavelet tour of signal processing. Academic, New York
89. Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Ecology 86:2497–2504
90. Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
91. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteor Soc 79:61–78
92. Grenfell BT, Bjørnstad ON, Kappey J (2001) Travelling waves and spatial hierarchies in measles epidemics. Nature 414:716–723
93. Cazelles B, Chavez M, De Magny GC, Guegan JF, Hales S (2007) Time-dependent spectral analysis of epidemiological time-series with wavelets. J R Soc Interface 4(15):625–636
94. Falkowski MJ, Smith AMS, Hudak AT et al (2006) Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. Can J Remote Sensing 32:153–161
95. Dray S, Legendre P, Peres-Neto PR (2006) Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). Ecol Model 196:483–493
96. Griffith DA, Peres-Neto PR (2006) Spatial modelling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87:2603–2613
97. Jombart T, Dray S, Dufour AB (2009) Finding essential scales of spatial variation in ecological data: a multivariate approach. Ecography 32:161–168
98. Jacquez GM, Maruca S, Fortin M-J (2000) From fields to objects: a review of geographic boundary analysis. J Geograph Syst 2:221–241
99. Fortin M-J, Drapeau P (1995) Delineation of ecological boundaries: comparison of approaches and significance tests. Oikos 72:323–332
100. Fortin M-J, Keitt TH, Maurer B et al (2005) Species ranges and distributional limits: pattern analysis and statistical issues. Oikos 108:7–17
101. McIntire EJB, Fortin M-J (2006) Structure and function of wildfire and mountain pine bettle forest boundaries. Ecography 29:309–318
102. Fortin M-J (1994) Edge detection algorithms for two-dimensional ecological data. Ecology 75:956–965
103. Oden NL, Sokal RR, Fortin M-J, Goebl H (1993) Categorical wombling: detecting regions of significant change in spatially located categorical variables. Geogr Anal 25:315–336

104. Fortin MJ, Drapeau P, Jacquez GM (1996) Quantification of the spatial co-occurrences of ecological boundaries. Oikos 77:51–60
105. St-Louis V, Fortin M-J, Desrochers A (2004) Association between microhabitat and territory boundaries of two forest songbirds. Landscape Ecol 19:591–601
106. Sadahiro Y, Umemura M (2002) A computational approach for the analysis of changes in polygon distributions. J Geograph Syst 3:137–154
107. Peters DPC, Bestelmeyer BT, Turner MG (2007) Cross-scale interactions and changing pattern-process relationships: consequences for system dynamics. Ecosystems 10:790–796
108. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York
109. Percival and Walden (2000) Wavelet methods for time series analysis. Cambridge, Cambridge University Press

Chapter 8
Ecosystems, Adaptive Management

Craig R. Allen, Joseph J. Fontaine, and Ahjond S. Garmestani

Glossary

Adaptive governance	Institutional and political frameworks designed to adapt to changing relationships between society and ecosystems, institutional frameworks that enable adaptive management, and the facilitation of learning from adaptive management to policy.
Adaptive management	A systematic process of natural resource management whereby management actions are treated as experiments to increase learning and improve subsequent management.
Natural resource management	The management of natural resources including land, water, plants, and animals to meet societal goals, including conservation and exploitation.
Resilience	The capacity of a system to absorb disturbance without altering states (undergoing a regime shift); a measure of the amount of disturbance a system can tolerate before collapsing.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

C.R. Allen (✉) • J.J. Fontaine
U.S. Geological Survey, Nebraska Cooperative Fish and Wildlife Research Unit,
School of Natural Resources, University of Nebraska, Lincoln, NE, USA
e-mail: allencr@unl.edu

A.S. Garmestani
U.S. Environmental Protection Agency, National Risk Management
Research Laboratory, Cincinnati, OH, USA

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_8,

Structured decision making	A general term for a framework of analysis of problems to reach decisions based on evidence to meet stated goals.

Definition of Adaptive Management

Adaptive management is an approach to natural resource management that emphasizes learning through management based upon the philosophy that knowledge is incomplete and much of what is thought to be known is actually wrong, but despite uncertainty, managers and policymakers must act [1]. Although the concept of adaptive management has resonated with resource management scientists and practitioners following its formal introduction in 1978 [2], it has and continues to remain little practiced and much misunderstood. Misunderstanding is largely based upon the belief that adaptive management is what management has always been, a trial and error attempt to improve management outcomes. But unlike a trial and error approach, adaptive management has explicit structure, including a careful elucidation of goals, identification of alternative management objectives and hypotheses of causation, and procedures for the collection of data followed by evaluation and reiteration. Since its initial introduction and description, adaptive management has been hailed as a solution to endless trial and error approaches to complex natural resource management challenges and recently, it has become increasingly referenced under various forms (please refer to following sections) (Fig. 8.1). Regardless of the particular definition of adaptive management used, and there are many, adaptive management emphasizes learning and subsequent adaptation of management based upon that learning. The process is iterative, and serves to reduce uncertainty, build knowledge, and improve management over time in a goal-oriented and structured process. However, adaptive management is not a panacea for the navigation of "wicked problems" [3, 4] as it does not produce easy answers, and is appropriate in only a subset of natural resource management problems where both uncertainty and controllability are high (Fig. 8.2) [5]. Where uncertainty is high but controllability is low, scenarios are a more appropriate approach. Adaptive management is a poor fit for solving problems of intricate complexity, high external influences, long time spans, high structural uncertainty, and with low confidence in assessments [5] (e.g., climate change). However, even in such situations, adaptive management may be the preferred alternative, and can be utilized to resolve or reduce structural uncertainty.

Clearly, adaptive management has matured, but it has also reached a crossroads. Its application is now common to a variety of complex resource management issues, and while practitioners and scientists have developed adaptive management and structured decision-making techniques, and mathematicians have developed approaches to reducing the uncertainties encountered in resource management, there continues to be misapplication of the method, and misunderstanding of its purpose.

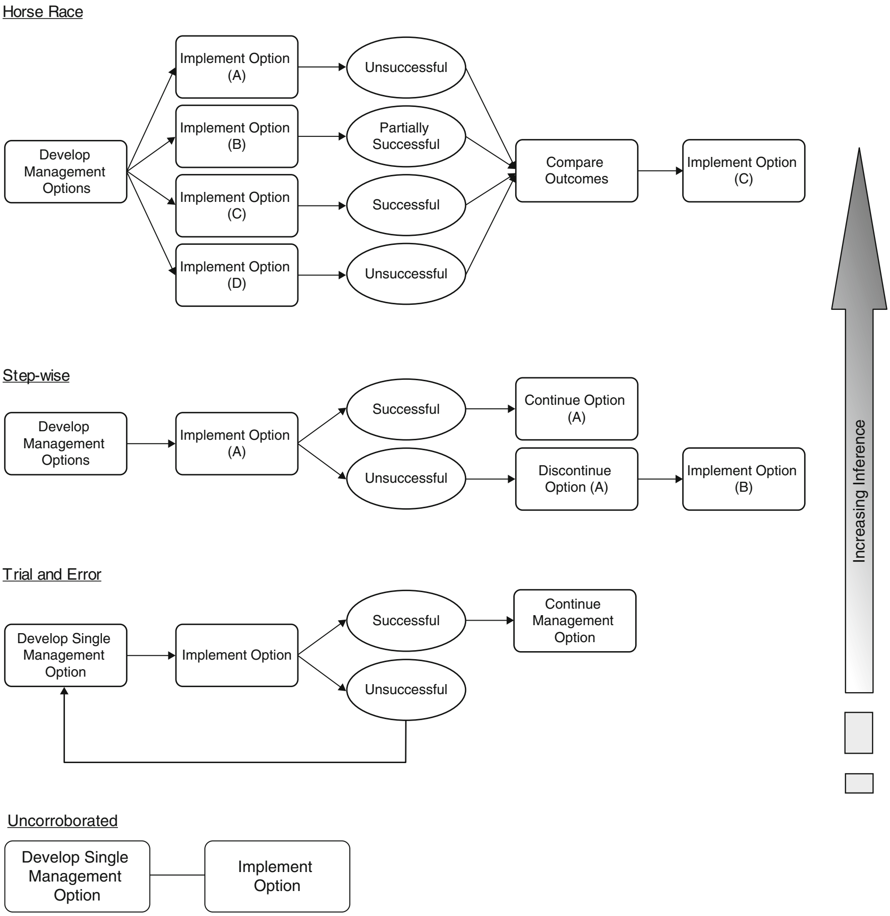


Fig. 8.1 Generalization of the different approaches to natural resource management

Introduction

Adaptive management of natural resources did not spontaneously appear, but represents an evolving approach to natural resource management in particular, and structured decision making in general. Founded in the decision approaches of other fields [6] including business [7], experimental science [8], systems theory [9], and industrial ecology [10], the first reference to adaptive management philosophies in natural resource management may be traced back to Beverton and Holt [11] in fisheries management, though the term "adaptive management" was yet to be used (reviewed in [6]). The term "adaptive management" would not become a common vernacular until C.S. Holling, widely recognized as the "father" of adaptive management, produced his edited volume on the subject "Adaptive

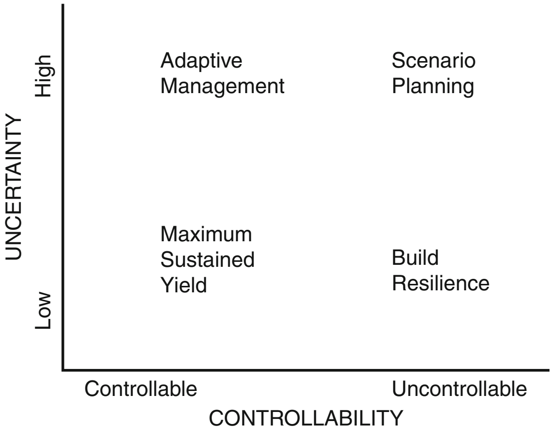


Fig. 8.2 Adaptive management and scenarios are complementary approaches to understanding complex systems. Adaptive management functions best when both uncertainty and controllability are high, which means the potential for learning is high, and the system can be manipulated (Adapted from [60])

Environmental Assessment and Management" in 1978 [2]. The work was spawned by the experiences of Holling and colleagues at the University of British Columbia following from the development of resilience theory [12]. The concept of resilience, predicated upon the existence of more than one alternative stable state for ecosystems, had several ramifications. For one, it meant that managers should be very careful not to exceed a threshold that might change the state of the system being managed, and the location of those thresholds is unknown. Second, for ecological systems in a favorable state, management should focus on maintaining that state, and its resilience. Adaptive management, then, was a method to probe the dynamics and resilience of systems while continuing with "management," whereby management experiments were developed to enhance learning and reduce uncertainty, in a fail-safe manner. According to Holling (http://www.resalliance.org/2561.php):

> The resilience research led us to mobilize a series of studies of large scale ecosystems subject to management- terrestrial, fresh water and marine. All this was done with the key scientists and, in some cases, policy people who "owned" the systems and the data. So the process encouraged two major advances. One advance developed a sequence of workshop techniques so that we could work with experts to develop alternative explanatory models and suggestive policies. We learned an immense amount from the first experiment. That focused on the beautiful Gulf Islands, an archipelago off the coast of Vancouver. We chose to develop a recreational land simulation of recreational property. I knew little about speculation, but we made up a marvelous scheme that used the predation equations as the foundation- the land of various classes were the "prey," speculators were the "predators" and a highest bidder auction cleared the market each year. The equations were modifications of the general predation equations. The predictions were astonishingly effective and persisted so for at least a decade. As much as anything, it reinforced the earlier conclusion that these equations were powerful and general. But the important conclusion concerned the workshop process and the people.

Eventually Carl Walters [1] built upon Holling's foundational contribution [12] and further developed the ideas, especially in the realm of mathematical modeling. Whereas Holling's original emphasis was in bridging the gap between science and practice, Walters emphasized treating management activities as designed experiments meant to reduce uncertainty. Both scientists sought an approach that allowed resource management and exploitation to continue while explicitly embracing uncertainties and seeking to reduce them through management. Walters [1] described the process of adaptive management as beginning "with the central tenet that management involves a continual learning process that cannot conveniently be separated into functions like research and ongoing regulatory activities, and probably never converges to a state of blissful equilibrium involving full knowledge and optimum productivity." He characterized adaptive management as the process of defining and bounding the management problem, identifying and representing what is known through models of dynamics that identify assumptions and predictions so experience can further learning, identifying possible sources of uncertainty and identifying alternate hypotheses, and finally the design of policies to allow continued resource management or production while enhancing learning.

A key focus of adaptive management is the identification and reduction, where possible, of uncertainty. Uncertainty is reduced through management experiments which enhance learning. Williams [6] describes four critical sources of uncertainty:

1. *Environmental variation* is often the most common source of uncertainty, and is largely uncontrollable. It may have a dominating influence on natural resource systems, through such factors as random variability in climate.
2. *Partial observability* refers to uncertainty about resource status. An example of this is the sampling variation that arises in resource monitoring.
3. *Partial controllability* arises when indirect means (e.g., regulations) are used to implement an action (e.g., setting a harvest rate), and it can lead to the misrepresentation of management interventions and thus to an inadequate accounting of their influence on resource behavior.
4. *Structural or process uncertainty* arises from a lack of understanding or agreement regarding the structure of biological and ecological relationships that drive resource dynamics.

Adaptive Management Today

Adaptive management has been referenced either implicitly [11] or explicitly [2, 13] for more than 50 years, but despite an illustrious theoretical history, there has remained imperfect realization of adaptive management in real world natural resource management decisions. The limited implementation of adaptive management stems from three fundamental problems: (1) a lack of clarity in definition and approach, (2) a paucity of success stories upon which to build [14–18], and (3) management, policy, and funding paradigms that favor reactive rather than proactive approaches to natural resource management [19, 20]. Each of these challenges has slowed the development of adaptive management as a paradigm for natural resource management and resulted in incomplete, inefficient, and even inappropriate implementation of adaptive management.

Although semantic arguments may seem the realm of ivory-towered professors, inconsistent and even contradictory approaches and definitions of adaptive management have resulted in confusion and limited the ability of management organizations to develop consistent and repeatable comprehensive adaptive management programs. Ironically, the confusion over the term "adaptive management" may stem from the flexibility inherent in the approach which has resulted in multiple interpretations of "adaptive management" that fall upon a continuum of complexity and *a priori* design, starting from the simple (e.g., "learning by doing") and progressing to the more explicit (e.g., "a rigorous process that should include sound planning and experimental design with a systematic evaluation process that links monitoring to management") [2, 21, 22]. Obviously, there is a clear distinction in intent, investment, and success between approaches that propose to learn from prior management decisions and those that outline a concise feedback mechanism dependent upon sound scientific principles on which future management decisions will be made. The definition of "adaptive management" is further confused because one of the powerful attributes of adaptive management is the ability to simultaneously address multiple needs of managers, scientists, and stakeholders. The result has been published reports of adaptive management that emphasize definitions that focus on the needs of the authors and the ability of adaptive management to meet those needs (e.g., experimentation [14], uncertainty [23], changing management actions [24], monitoring [25], and stakeholder involvement [26]).

Despite the challenges in defining adaptive management, momentum and interest in the subject and its application continue to grow. The recent development by the United States Department of Interior of an adaptive management technical guide (http://www.doi.gov/initiatives/AdaptiveManagement/TechGuide.pdf) and the policies developed around this manual to:

> Incorporate adaptive management principles, as appropriate, into policies, plans, guidance, agreements, and other instruments for the management of resources under the Department's jurisdiction. – Department of Interior Manual (522 DM 1)

are an indication of the growing movement in natural resource management toward taking a more proactive role in management decisions. Unfortunately, this movement has little to build upon with one clear exception, the U.S. Fish and Wildlife Service (USFWS) Adaptive Harvest Management Plan (AHM) for mid-continent mallards. Worldwide, AHM is one of the few successful efforts to apply the principles of adaptive management and demonstrate how to successfully manage natural resources by improving the understanding of natural systems through management actions. The adaptive management processes of AHM have greatly improved the understanding of the harvest potential of waterfowl populations, the ability of managers to regulate harvest, and the importance of monitoring and assessment programs to support the decision-making process.

So why has AHM succeeded while so many other attempts to implement adaptive management have stalled? First, AHM developed a clear and concise objective: maximize long-term waterfowl harvest while ensuring long-term viability of waterfowl populations. The development and agreement by stakeholders to a concise set of fundamental objectives is paramount to ensuring the success of any adaptive management program. Failure to agree upon fundamental objectives and unwarranted attempts to alter objectives will ensure any attempt to manage, whether adaptive or not, will fail. The second key to the AHM success was due to simultaneous support for management, research, and monitoring. Waterfowl research and management in North America are nearly unequaled by almost any natural resource management program in terms of history, scope, and investment [27]. The enormity of historical and current data and the availability of resources for researchers and managers to utilize that data have facilitated the development of innumerable research and management activities all of which have fed back into the AHM process. In addition, the AHM program has arguably one of the most comprehensive monitoring programs for any ecological system currently under study. The combination of well-supported management, research, and monitoring programs has resulted in a clear reduction in the uncertainty of how waterfowl populations respond to management and enabled managers and policy makers to more effectively meet their stated objectives. Unfortunately, too often, attempts to implement adaptive management fail to address all of the requirements. In particular, resources for monitoring and research are often undervalued with the resultant outcome being a series of management actions with no understanding of their implications.

The final key to the success of AHM has been the ability to implement management and policy decisions based on the best information available. In many historical and current attempts to implement adaptive management, the regulatory body charged with implementation of management recommendations either is unable, or worse, is unwilling to implement actions proposed by the outcome of the adaptive management process. The body in charge of regulatory control is too often a stakeholder in the process of adaptive management with an agenda independent of regulating the resource alone. There may even be, and often are, several regulatory agencies controlling resources, each an independent stakeholder, each with an independent agenda. Such a situation can make implementation of a management recommendation challenging, especially if it contradicts long-standing dogma. Consider for example, the management of Glen Canyon Dam and the waters of the Colorado River. Heralded by Congress as an adaptive management success story, the Colorado River Adaptive Management Program has fallen short of success because despite 13 years of work, the ecological status of the Colorado River and the conflict inherent to the development of an adaptive management program continue to worsen [28]. This is because the regulatory agency that controls the flow of water throughout the Colorado River Basin, the Bureau of Reclamation, is also one of the major stakeholders in the adaptive management process with an agenda (water storage) that conflicts with several other stakeholders and regulatory agencies that manage people and wildlife along the Colorado River (e.g., California Department of Water Resources, Mexican National Water Commission, USFWS). In contrast to the management of the

Colorado River, there is a single centralized regulatory body governing waterfowl harvest in the United States (USFWS), and although there are many stakeholders that play a role in setting harvest management regulations, ultimately, decisions are made by the USFWS. Equally important, the interests of the USFWS parallel those of the other stakeholders. For the Colorado River, stakeholder interests are almost directly at odds. So from these examples is one to conclude that adaptive management is an unattainable mandate for the management of resources where various stakeholders and regulators are at odds? No, implementation of adaptive management is appropriate in both examples, possibly even more so for the management of the Colorado River. What the Colorado River example highlights is the importance of collaboration, the benefits of a single or superregulatory body, and the need to agree upon *a priori* objectives that guide long-term management decisions despite short-term political, societal, economic, or even environmental impacts.

Structured Decision Making

A key component of any management approach, whether it is adaptive or not, is deciding on the objectives, goals, and ultimately management options that may best achieve the desired goals (Fig. 8.3). Unfortunately, as with many decisions, deciding upon a proper set of objectives and the means to reach those objectives can prove challenging. Resource management decisions are further complicated because social-ecological systems are complex (e.g., multiple objectives and stakeholders, overlapping jurisdictions, short- and long-term effects) and are characterized by a high degree of uncertainty (e.g., appropriate management action or monitoring protocols, future economic or ecological conditions) and therefore present decision makers with challenging judgments (e.g., predicted consequences of proposed alternatives, value-based judgments about priorities, preferences, and risk tolerances) often under enormous pressure (economic, environmental, social, and political) and with limited resources to ensure success. The resulting outcome of such conditions too often leads to management paralysis, or continuation of the status quo, as managers and policy makers become overwhelmed by the process of the decision and lose track of the desired social-ecological conditions they are charged with achieving. Indeed, the process of resource management can be arduous and even controversial, particularly if there are a variety of stakeholders vying to push the agenda. Fortunately, there are methods to overcome these pitfalls and maximize the potential for success.

One method to overcome management paralysis and mediate multiple stakeholder interests is structured decision making. Borrowed from the sociological fields, structured decision making is an organized approach to identify and evaluate alternative resource management options by engaging stakeholders, experts, and decision makers in the decision process and addressing the complexity and uncertainty inherent in resource management in a proactive and transparent manner.

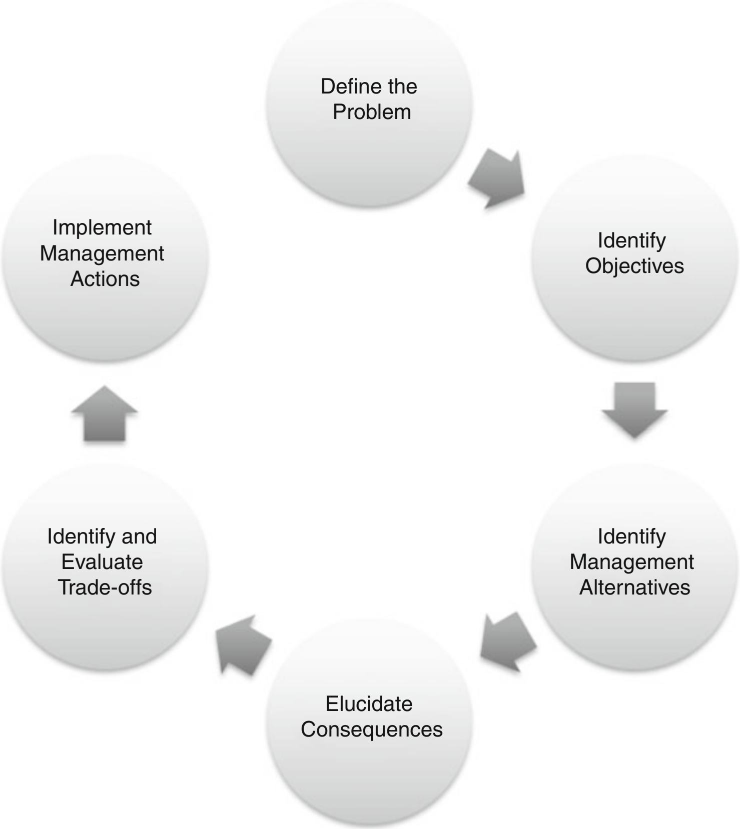


Fig. 8.3 The minimum steps necessary to implement a structured decision-making process: More complex integration of individual steps may be necessary if future steps clarify the process or if the decision is iterative over time

Structured decision making uses a simple set of steps (Fig. 8.3) to evaluate a problem and integrate planning, analysis, and management into a transparent process that provides a roadmap focused on achieving the fundamental objectives of the program. It differs somewhat from "active" adaptive management in that it does not emphasize replicated management experiments (Fig. 8.4). Central to the success of the structured decision making process is the requirement to clearly articulate fundamental objectives, explicitly acknowledge uncertainty, and respond transparently to all stakeholders' interests in the decision process. The conceptual simplicity inherent in structured decision making makes the process useful for all decisions from minor decisions to complex problems involving multiple stakeholders.

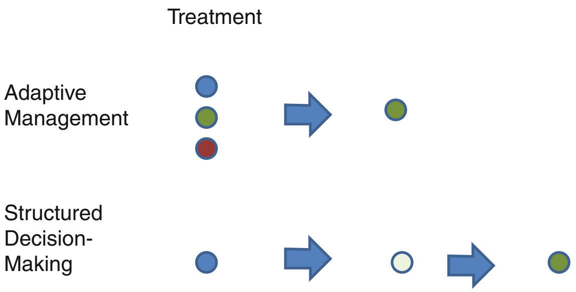


Fig. 8.4 Structured decision making and adaptive management differ somewhat, especially in that active adaptive management emphasizes the utilization of multiple replicated management experiments. As such, learning may be faster when such experiments are possible. However, adaptive management and structured decision making are terms often used interchangeably

Structured Decision Making Steps

1. *Define the Problem* – The first step in a structured decision making process is a clear and concise evaluation and articulation of the problem being addressed and the motivation underlying the need to address the problem. Although identifying the problem may seem self-evident, failure to clearly articulate the problem to all stakeholders and subsequent agreement by stakeholders as to the nature of the problem is often cited as the primary reason management and policy actions fail, or worse, face future litigation. To facilitate this process, decision makers need to ask:

   (a) What specific decision(s) have to be made?
   (b) What is scope of the decision (e.g., geographic, temporal)?
   (c) Will the decision be iterated over time?
   (d) What are the constraints within which the decision will be made (e.g., logistical, ecological, legal, temporal, financial)?
   (e) What stakeholders should be involved in the decision process and what are their respective roles?

2. *Identify the Objectives* – The centerpiece of the structured decision making process is a set of clearly elucidated objectives. Together they define the "why do we care" about the decision and thereby facilitate the search for alternatives, and become the metric for comparing and evaluating management outcomes. When defining objectives, there are many considerations to ensure that decision makers can adequately evaluate alternatives. Ideally, objectives are stated in quantitative terms that relate to parameters that can be measured and thus evaluated. More importantly, objectives are meant to focus efforts on the importance of the decision in a consistent and transparent manner that exposes key trade-offs and uncertainties so decision makers can generate creative and proactive alternatives. Objectives should be complete, controllable, concise, measurable, and understandable [29]. To achieve this end requires "brainstorming" with

stakeholders to identify what is important about the decision at hand. The outcome of such an effort may produce a wide variety and often extensive list of objectives that will need to be simplified to focus on things that matter and the direction they need to move (e.g., maximize deer harvest or minimize erosion). It is important to note, that unlike goals or targets, objectives do not have specific quantitative outcomes (e.g., 50% increase), but are meant to define the preferred ends and the direction of change to meet that ends.

Once a list of objectives has been defined, it is important to separate the objectives into fundamental objectives (which reflect the ultimate goals) and means objectives (which are ways of achieving the ends) to ensure that management actions really effect the defined problem. For example, "maximize sandbars" may be an important objective for the management of a river like the Missouri or Platte, but if the river system is being managed for wildlife, sandbars are primarily important because they increase breeding habitat for threatened and endangered terns and plovers. "Maximize sandbars" is thus a means objective toward reaching the fundamental objective of "maximize tern and plover population size." Clearly, there are other means objectives that would also facilitate this fundamental objective (e.g., minimize nest predation, maximized food availability, etc.). The benefit of the process of distinguishing objectives is that the identification of means objectives can help lead to alternative management actions (e.g., build sandbars, release reservoir water), while the identification of fundamental objectives gives a basis for evaluating and comparing alternatives (annual tern and plover population size). Keep in mind, however, that the status of fundamental or means is not an innate quality of an objective, but rather is highly context dependent. Thus, what was a means objective for one decision, in the example "maximize sandbars," may be a fundamental objective for another if the decision problems shifts from say "wildlife management" to "aesthetics" or "flow."

After developing a careful list of objectives, it can be useful to develop a hierarchy or means-ends diagram to group similar objectives and clarify the links and relationships between means and fundamental objectives. An objectives hierarchy can help clarify the context of each fundamental objective by identifying all the important elements that are affected by the decision process and demonstrate to stakeholders the importance of all objectives even those that are not "fundamental objectives."

3. *Identify Management Alternatives* – Management success is only as likely as the creativity and diversity of possible management alternatives. Unfortunately, management paralysis, "pet" management actions, and staying with the status quo too often limit managers and policy makers to few options and thereby impede management success. The process of identifying management alternatives, like the process of identifying objectives, starts with brainstorming. Identifying alternative management actions is a process that should be addressed iteratively, as knowledge of best practices and the creativity to develop novel ideas should not be expected to develop instantaneously. The key is bringing the "right" people together. It is important to have a group with a set of

interdisciplinary backgrounds that represent the larger decision to ensure that the needs of stakeholders are not overlooked. This is not to say that the stakeholders involved in identifying alternative management actions are the same as the larger stakeholder group, usually they are not. This is primarily due to the technical knowledge necessary to present plausible alternatives. Still there are opportunities where the benefit of being naive may present novel actions that might not otherwise be considered.

The brainstorming process should begin by identifying alternatives for individual objectives, but always be looking for opportunities when one action may fulfill the needs of multiple objectives. Identifying alternatives also means being mindful of those actions that must be done (e.g., standing policy), constraints (real or perceived) and potential trade-offs between objectives and various management actions. In developing alternatives, it is important that the "brainstorming" process focus on developing high-quality management actions that are: (1) explicitly designed to address the outlined objectives, (2) technically sound in that they build on the best known practices, (3) concise yet comprehensive enough to include the technical understanding for implementation, (4) designed to expose trade-offs between the decision process by having mutually exclusive strategies, and (5) developed to achieve the greatest good for the stakeholders involved.

Once an extensive list of alternatives has been identified, it can be useful to group them into strategies or portfolios based on general similarities in what they aim to achieve. Sometimes these portfolios can represent the needs of specific stakeholder groups or specific conditions that could be achieved. For example, management actions on a river system may be grouped together into portfolios that meet the needs of sport-fishery, endangered species, or irrigation; alternatively, they may be grouped based on their ability to return the river to 50%, 75%, or 95% of historical flows. Both methods have merit, the first in that it is generally clear to the stakeholders what objectives are being met and then where trade-offs must be considered, and the second in that the inherent interests of any particular group are not the driving factor and thus the process can be less contentious.

4. *Elucidate Consequences* – The list of alternative management actions is only effective if it creates an opportunity to evaluate and compare actions in light of the objectives before implementation. It is important to realize that the process of identifying management consequences is not a value judgment, but an analytical assessment of the most likely outcome of the action(s). Using the best scientific knowledge available, this process is a modeling exercise focused on predicting the likely outcomes of each alternative and thus the likelihood that each achieves the desired objective. Depending upon our knowledge of the system, this process can be highly quantitative where extensive data are modeled and probabilities assigned to each outcome or as is often the case, if little or nothing is known about the system, this process can depend heavily on expert opinion or comparisons to similar systems. In both cases, there is a degree of uncertainty associated with predicted outcomes as well as the parameters

included in the modeling process. Indeed, because system function is rarely precisely understood, the effects of management actions are never certain and the future states are unknown, decisions are almost always made in the face of uncertainty. Uncertainty can make differentiating among alternatives difficult, but because uncertainty is an inherent part of the decision process, it must not be ignored. It is important that uncertainty be confronted throughout the decision process and that the uncertainties are identified and the possible impacts on the system and the ability to achieve stated objectives documented.

Once the modeling process has predicted the likely outcomes of each management action and the corresponding ability to address each objective, the next step is to develop a consequence table. The purpose of a consequence table is to produce a visual summary of the consequences of each potential management action on each of the objectives in a table or matrix. A consequence table can take a variety of forms, from a simple rating system (e.g., consumer report 5-star rating) to a complex table with specific probabilities of outcomes and subsequent likelihoods of achieving each objective. Independent of the complexity of the underlying models that populate the matrix, the purpose of the consequence table is to ease and facilitate direct comparison of each management actions' ability to achieve each objective.

5. *Identify and Evaluate Tradeoffs* – Ideally the structured decision making process would lead to a clear management alternative that achieves the objectives of all interested parties; unfortunately, this is rarely the case. Generally, the process of developing a consequence table will clearly elucidate which options are the least likely to be effective, but if there are multiple stakeholders and thus multiple objectives, most decisions will require a trade-off between the ability of the remaining options to achieve each objective. The process of identifying where these trade-offs arise is analytical, but the decision process itself is highly value laden and thus dependent upon stakeholders. In most complex decisions, this will involve stakeholders choosing between less-than-perfect alternatives. There are a variety of methods to facilitate highly value-laden decisions by weighing options based on the values of the stakeholders and then comparing alternatives to find the "best" compromise solutions. However, trade-offs are real and it is unlikely that all parties will be totally satisfied with the eventual outcome. Indeed, although consensus is ideal, it is not necessary and is often unachievable; however, the benefit of the structured decision-making process is that even if there is disagreement, the process makes the disagreement transparent and enables stakeholders to re-evaluate using new knowledge and/or perspectives.
6. *Implement Management Action* – The final step in the structured decision-making process is implementation. Although this may always seem to be the desired outcome of a decision process, unfortunately, social and political pressures to reach "perfection" often impede implementation and leave decisions in a continuous state of inaction. To ensure success, managers, policy makers, and stakeholders must work together to move through the decision process in a timely manner to ensure action can be taken. Failure to take action is a decision, whether it is made passively or actively.

Participatory Active Adaptive Management

Panarchy is a useful model for characterizing ecological systems and the formal institutions that manage these systems [30]. One of the most critical aspects in the panarchy appears to be a bridging organization that can monitor the status of the social-ecological system, and manifest rapid change, if conditions are deteriorating [31]. Monitoring will allow for management to set new target levels, and modify policy to reach those target levels, as new information is generated on scale-specific system attributes [32]. In order for management entities operating at discrete scales to improve communication channels and create opportunities for collaboration, intermediate level entities may serve to facilitate these cross-scale linkages. Bridging organizations have the capacity to fulfill this role and organize cooperation between stakeholders across scales [33], but to do so successfully, one must formulate strategies, coordinate joint action, address uncertainty, and link diverse stakeholders in a world of increasing complexity. Brown [33] investigated bridging organizations from across the world, and from a variety of scopes (e.g., regional economic policy in the USA; small-scale irrigation projects in Indonesia; agricultural productivity in Zimbabwe) found that bridging organizations are independent of stakeholders in a social-ecological system, which allows them to negotiate with stakeholders and advocate multiple positions. This unique role in the management of social-ecological systems affords bridging organizations the capacity to catalyze the formation of policies that are flexible and reflective of the panarchy of ecosystems and institutions [33]. In addition, bridging organizations have the capacity to reduce transaction costs, and provide a mechanism to enforce adherence to desired policies, despite their lack of regulatory authority [34].

Examples of bridging organizations include: (1) assessment teams, which are made up of actors across sectors in a social-ecological system; (2) nongovernmental organizations, which create an arena for trust-building, learning, conflict resolution, and adaptive co-management; and (3) the scientific community, which acts as a “watchdog,” as well as a facilitator, for adaptive management. For purposes of environmental management, an example of a successful bridging organization is that of Ekomuseum Kristianstads Vattenrike (EKV), a small, municipal organization that facilitated progressive ecosystem management in southern Sweden [34]. EKV was tasked with managing water resources at a regional scale in Sweden, and was successful largely because it employed organizational flexibility that allowed for EKV to respond quickly to “surprise.” This was achieved through leadership, a core interdisciplinary staff, and the facilitation of connections between individuals and organizations (i.e., the panarchy of institutions) in the social-ecological system. EKV was able to improve the social capacity to respond to “surprises” and create the trust necessary to push the social-ecological system toward improved adaptive management of resources.

The formal management institutions in place are likely to persist barring a large-scale perturbation to social-ecological systems. So, managers must operate within the limitations of these institutions, which complicates matters, but does not make the situation intractable. One possible option for improving environmental management, as highlighted in this section, appears to be in developing bridging organizations that catalyze cross-scale communication across the panarchy of institutions and ecosystems, and explicit recognition of the underlying cross-scale structure and nonlinear interactions of these linked systems, by both policy and policy makers. The lack of communication and cooperation between institutions at even small scales further illuminates that bridging organizations may help bring about effective management of natural resources at multiple scales [35]. Thus, bridging organizations should act as mini think tanks that facilitate communication between institutions, incubate new ideas for environmental management, and provide a forum for coming to agreements on contentious issues [36].

Bridging organizations play a critical role in facilitating adaptive comanagement and governance, and are essential to managing for resilience in social-ecological systems [37]. Perception of a particular policy can play a significant role in whether it is accepted by critical stakeholders in a social-ecological system [38]. Engaging stakeholders, implementing change at a suitable rate, and providing outreach to keep the public informed are all important for new environmental policy to be perceived of as positive and for a successful transition to a new policy regime [38]. This environmental management framework, which incorporates panarchy, adaptive management, and bridging organizations, could serve as one scenario in the suite of policy options for actualizing sustainability [30].

Adaptive Governance

Administrative agencies typically change incrementally [39], and as such, changes in policy are small because there is not enough information to make large overhauls of organization policy. Standard operating procedures are another mechanism that contributes to organizational inertia, as they slow the bureaucratic process [40]. Further, the lack of institutions matched to the appropriate scale is a significant barrier for sound environmental management [41]. Within this context, adaptive governance can help with this scale mismatch via collaboration of a diverse set of stakeholders at multiple scales [42]. Adaptive governance is a form of governance that incorporates formal institutions, informal groups/networks, and individuals at multiple scales for purposes of collaborative environmental management [43]. Bridging organizations, enabling legislation and government policies can also contribute to the success of an adaptive governance framework; governance creates a vision and management actualizes the vision [43].

Adaptive governance works via sharing of management power and responsibilities, and promotes a collaborative, participatory process, but is dependent upon adaptive comanagement, and adaptive comanagement is dependent upon social networks for success. Social networks have the capacity to allow for development of new ideas, to facilitate communication between entities, and to create the flexibility necessary for the interplay of the fluid (ecological systems) and the rigid (institutions) to be successful for environmental management [43]. Leadership has been well established as a critical factor in facilitating good environmental management. Leaders develop and facilitate a vision for environmental management, incorporating local knowledge and information from social networks [43].

Olsson et al. [44] studied adaptive comanagement in Sweden and Canada and concluded that this form of management of ecological systems was most effective when there was: leadership with vision for the system of interest; legislation that created the environment for adaptive management; funds for adaptive management; monitoring of the ecological system; information flow (i.e., cross-scale linkages); combination of a variety of sources knowledge; and venue for collaboration. Olsson et al. [44] contend that these factors are critical to building resilience in social-ecological systems, as they help to protect the system from the failure of management decisions under uncertainty (i.e., imperfect information). Further, they assert that adaptive comanagement is necessary to facilitate adaptive governance. In turn, adaptive governance is facilitated by informal networks and leadership, which creates the capacity for development of novel ideas for environmental management [43]. These informal networks have the capacity to generate political, financial, and legal support for novel environmental management [43]. Further, adaptive governance is dependent upon polycentric institutions that are redundant (e.g., scale-specific) and are quasi-autonomous [45]. Olsson et al. [45] compared five case studies from around the world and concluded that in order for a social-ecological system to transition to adaptive governance, it must undergo a preparation and a transformation phase, linked by a window of opportunity.

In a well-cited example (Kristianstads Vattenrike) from Sweden, Olsson et al. [45] report the transition to adaptive governance was preceded by the development of a social network of parties interested in the management of the social-ecological system. The network consisted of members from local groups (environmental groups, farmers' associations), local government (municipality of Kristianstad, the County Administrative Board), and national scale (World Wildlife Fund, National Museum of Natural History, National Research Council). In case studies that have not resulted in a successful transition to adaptive governance, the social networks needed to help facilitate the transition were not well developed, and this hindered the changes needed for good environmental management [45].

The role of leadership has also been cited as critical to a transition to adaptive governance, and Olsson et al. [45] provide an example of leadership from Kristianstads Vattenrike. A key individual acted as a catalyst to social network formation, setting the research agenda, and mobilizing support at multiple scales for "new" environmental management. Critical to setting an agenda is defining how an issue becomes perceived as a "public problem because if most individuals accept a particular condition, negative feedback works to maintain public opinion in that particular regime" [46]. However, if the individuals in the regime develop a "critical mass" of distaste for a particular issue, public opinion can cross a threshold and reorganize into an alternative regime. Importantly, interest groups, the media, and other agents can have an effect on agenda setting and creating the "climate" necessary for a shift in public opinion [46]. There are critical roles to be played by individual actors in shifting policy from one regime to an alternate regime. For instance, social networkers that share information freely; individuals that have numerous, diverse connections; and individuals with powerful ability to persuade play key roles in policy change [47]. These individuals can interact to create the conditions necessary for regime shifts in public policy. In particular, the director of a municipal organization (Ekomuseum Kristianstads Vattenrike) filled this leadership role and served as a bridging organization that also was a significant factor in the transition to adaptive governance [45]. The leadership needed to foster a transition to adaptive governance is not necessarily the work of one individual, but rather is often encompassed by several individuals and entities [48].

There are two types of policy windows: a problem-driven window and a politically driven window [49]. A problem-driven window opens when a policymaker believes that a policy is necessary for a specific issue. A politically driven window is driven by a particular theme adopted by a policymaker, in which the policymaker looks for problems that fit within the theme. Significant changes in policy occur when conditions (e.g., problems, solutions, and politics) converge at the same time, which creates the window of opportunity for change [49]. In the Kristianstads Vattenrike example, social and ecological change at one scale triggered cross-scale effects which resulted in a window of opportunity for the transition to adaptive governance [45]. In adaptive governance, decision making is not top-down but rather emerges from outreach and group meetings with stakeholders [50]. In order for adaptive governance to be effective, the policy requires strong leadership, communication, and incorporation of uncertainty, which allows for adaptation to changing circumstances [50].

Adaptive Management and Law

Legal certainty is an aspect of law that does not mesh well with environmental unpredictability. One of the most significant barriers for managing linked social-ecological systems is that often the aspects of a society that make it free (e.g., certainty of law) are not in concert with ecological realities (e.g., multi-regimes, nonlinear systems, and responses) [51]. The certainty of law and institutional rigidity often limit experimentation that is necessary for adaptive management [30]. This point is critical, as some scholars contend that environmental governance of the commons can only succeed if rules evolve with the system of interest [41].

Ecosystem management has been applied within the outdated framework of the Endangered Species Act (ESA), but ecosystem management is best implemented via adaptive management [52]. In its current form, the ESA does not have the necessary flexibility in its regulatory language to effectively implement adaptive responses to changing environmental conditions [52]. The legal constraints upon adaptive management in the American system of law do not stop there. The fundamental constraint to adaptive management is the current state of administrative law [53]. As the law now stands, the procedural rules require a vast amount of work before an agency promulgates a rule or issues a permit [54]. This "pre-decision" activity allows for public input and prepares agencies for judicial review. Ruhl [54] contends that "agencies will find that interest groups and courts relentlessly will erode adaptive agency behavior, using all the tools conventional administrative law puts at their disposal." Having to operate in an atmosphere where each policy is evaluated on the "front-end," in anticipation of public and legal scrutiny, has squelched agencies' appetite for adaptive management.

US administrative law is a two-step process, in which the first step allows for public comment on draft documents and alternative options [55]. The second step is final agency action, which creates "certainty" to the process and makes the decision subject to judicial review. This process is based on the assumption that agencies have the capacity to predict the consequences of a "final agency action" [55]. Thus, there is a fundamental conflict between linear legal processes (i.e., administrative law) based on "stationarity," versus environmental management frameworks (i.e., adaptive management) based on the realization of dynamic systems characterized by "surprise" [55]. Given this inherent conflict, adaptive management may not be possible under the current administrative law framework [54].

The adversarial character of administrative law, combined with the need for certainty (e.g., procedural rules) in the larger realm of American law, is likely incompatible with adaptive management [56]. Thus, environmental law is at odds with science, as the certainty required for socio-political stability makes it very difficult to apply a novel approach to ecosystem management (e.g., adaptive management) that requires institutional flexibility. Thus, if adaptive management is necessary for good environmental management, environmental law must be "adapted" to fit with adaptive management [54]. Karkkainen [56] argues that administrative law should proceed on two trajectories: (1) a fixed rule track that will apply unless an agency can justify otherwise; and (2) an adaptive management track, where a new set of administrative law standards specific to adaptive management would hold precedence, in order to actualize adaptive management as a tool for environmental policy.

Thus, some in the law community argue that adaptive management is not possible under the current administrative law framework [54]. The National Environmental Policy Act (NEPA) may act as a barrier to implementation of adaptive management (sensu Holling) [57]. NEPA could possibly be modified to an iterative process that could accommodate adaptive management [57]. Ruhl [54] contends that adaptive management is necessary for good environmental management, which in turn means that environmental law must be "adapted" to fit with adaptive management.

In effect, administrative agencies in the USA do not conduct adaptive management as it was originally conceived [55]. Rather, agencies conduct adaptive management "lite," as the courts have provided some leeway to adaptive management projects, provided they have requirements that are legally enforceable [55]. The primary problem with adaptive management "lite" is that it does not measure up to the standards of adaptive management theory, nor does it hold up under the scrutiny of substantive and procedural law. Adaptive management (sensu Holling) is not likely until Congress provides more funding for adaptive management and clear standards for the adaptive management process [55].

Conclusions

The conceptual underpinnings for adaptive management are simple; there will always be inherent uncertainty and unpredictability in the dynamics and behavior of complex ecological systems as a result of nonlinear interactions among components and emergence, yet management decisions must still be made. The strength of adaptive management is in the recognition and confrontation of such uncertainty. Rather than ignore uncertainty, or use it to preclude management actions, adaptive management can foster resilience and flexibility to cope with an uncertain future, and develop safe-to-fail management approaches that acknowledge inevitable changes and surprises. Since its initial introduction, adaptive management has been hailed as a solution to endless trial and error approaches to complex natural resource management challenges. However, it does not produce easy answers, and it is appropriate in only a subset of natural resource management problems. Clearly adaptive management has great potential when applied appropriately.

Future Directions

Adaptive management is increasingly heralded as the future of natural resource management and has been adopted by many governmental and nongovernmental agencies. Institutions adopting adaptive management have utilized different definitions often focusing on a single strength of the process (i.e., experimentation, reducing uncertainty, involving stakeholders) and thus operationalize the practice uniquely. Some, like the U.S. Department of Interior, are highly focused on the decision process and the incorporation of structured decision making while others, such as the US Army Corps of Engineers, have embraced stakeholder involvement. Each approach has merit but adaptive management has failed to live up to its expectations [58]. The reasons for failure are many, and likely to be repeated, yet the great potential of adaptive management remains; unfortunately, it remains largely untapped. Translation of adaptive management approaches to “on-the-ground” natural resource managers is a critical step that has largely failed. Most natural resource managers are still unable to define adaptive management, let alone incorporate it into their normal management activities. The next decade will be critical: Will adaptive management remain in the domain of ivory towers, or will it become a tool for the trenches? Taking adaptive management to the practitioners will require the communication of adaptive management techniques in a clear, simple, and most importantly applicable manner. Currently, adaptive management fails because of an adherence to mathematical modeling above all else, its application to situations that are not conducive to replication or the measurement of success (e.g., large rivers such as the Missouri or the Colorado), and because adaptive management has not been adequately incorporated into natural resources management via appropriate legal mechanisms [59]. If the future of natural resource management is to be proactive and address the increasing uncertainties facing our world, adaptive approaches to resource management will require communication of the methodology and merits in a clear and simple manner.

Acknowledgments The Nebraska Cooperative Fish and Wildlife Research Unit is jointly supported by a cooperative agreement between the United States Geological Survey, the Nebraska Game and Parks Commission, the University of Nebraska−Lincoln, the United States Fish and Wildlife Service and the Wildlife Management Institute. Any use of trade names is for descriptive purposes only and does not imply endorsement by the US Government.

Bibliography

1. Walters CJ (1986) Adaptive management of renewable resources. McMillan, New York
2. Holling CS (ed) (1978) Adaptive environmental assessment and management. Wiley, Chichester
3. Rittel HWJ, Webber MM (1973) Dilemmas in a general theory of planning. Policy Sci 4:155–173
4. Ludwig D (2001) The era of management is over. Ecosystems 4:758–764
5. Gregory R, Ohlson D, Arvai J (2006) Deconstructing adaptive management: criteria for applications to environmental management. Ecol Appl 16:2411–2425
6. Williams BK (2010) Adaptive management of natural resources – framework and issues. J Environ Manage 92(5):1346–1353
7. Senge PM (1990) The fifth discipline: the art and practice of the learning organization. Currency Doubleday, New York
8. Popper KR (1968) The logic of scientific discovery, 2nd edn. Harper and Row, New York
9. Ashworth MJ (1982) Feedback design of systems with significant uncertainty. Research Studies Press, Chichester
10. Allenby BR, Richards DJ (1994) The greening of industrial ecosystems. National Academy Press, Washington, DC
11. Beverton RJH, Holt SJ (1957) On the dynamics of exploited fish populations. Her Majesty's Stationery Office, London
12. Holling CS (1973) Resilience and stability of ecological systems. Annu Rev Ecol Syst 4:1–23
13. Walters CJ, Hilborn R (1978) Ecological optimization and adaptive management. Annu Rev Ecol Syst 9:157–188
14. Lee KN (1993) Compass and gyroscope: integrating science and politics for the environment. Island Press, Washington, DC
15. McLain RJ, Lee RG (1996) Adaptive management: promises and pitfalls. Environ Manage 20:437–448
16. Lee KN (1999) Appraising adaptive management. Conserv Ecol 3(2):3. http://www.consecol.org/vol3/iss2/art3/
17. Moir WH, Block WM (2001) Adaptive management on public lands in the United States: commitment or rhetoric? Environ Manage 28:141–148
18. Walters C (2007) Is adaptive management helping to solve fisheries problems? Ambio 36:304–307
19. Ascher W (2001) Coping with complexity and organizational interests in natural resource management. Ecosystems 4:742–757
20. Schreiber ESG, Bearlsin AR, Nicol J, Todd CR (2004) Adaptive management: a synthesis of current understanding and effective application. Ecol Manage Restor 5:177–182

21. Wilhere GF (2002) Adaptive management in habitat conservation plans. Conserv Biol 16:20–29
22. Aldridge CL, Boyce MS, Baydack RK (2004) Adaptive management of prairie grouse: how do we get there? Wildl Soc Bull 32:92–103
23. Williams BK, Johnson FA (1995) Adaptive management and the regulation of waterfowl harvests. Wildl Soc Bull 23:430–436
24. Rauscher HM (1999) Ecosystem management decision support for federal forests in the United States: a review. For Ecol Manage 114:173–197
25. Bormann BT, Haynes RW, Martin JR (2007) Adaptive management of forest ecosystems: did some rubber hit the road. Bioscience 57:186–191
26. Norton BG (1995) Ecological integrity and social values: at what scale? Ecosyst Health 1:228–241
27. Hawkins AS, Hanson RC, Nelson HK, Reeves HM (eds) (1984) Flyways: pioneering waterfowl management in North America. U.S. Government Printing Office, Washington, DC
28. Susskind L, Camacho AE, Schenk T (2010) Collaborative planning and adaptive management in Glen Canyon: a cautionary tale. Columbia J Environ Law 35:1–54
29. McDaniels T (2000) Creating and using objectives for ecological risk assessment and management. Environ Sci Policy 3:299–304
30. Garmestani AS, Allen CR, Cabezas H (2009) Panarchy, adaptive management and governance: policy options for building resilience. Neb Law Rev 87:1036–1054
31. Kinzig A, Starrett D, Arrow K, Aniyar S, Bolin B, Dasgupta P, Ehrlich P, Folke C, Hanemann M, Heal G, Hoel M, Jansson A, Jansson BO, Kautsky N, Levin S, Lubchenco J, Maler KG, Pacala SW, Schneider SH, Siniscalco D, Walker B (2003) Coping with uncertainty: a call for a new science-policy forum. Ambio 32:330–335
32. Karkkainen BC (2002) Collaborative ecosystem governance: scale, complexity and dynamism. Virginia Environ Law J 21:189–244
33. Brown LD (1993) Development bridging organizations and strategic management for social change. Adv Strat Manag 9:381–405
34. Hahn T, Olsson P, Folke C, Johansson K (2006) Trust-building, knowledge generation and organizational innovations: the role of a bridging organization for adaptive co-management of a wetland landscape around Kristianstad, Sweden. Hum Ecol 34:573–592
35. Roy AH, Wenger SJ, Fletcher TD, Walsh CJ, Ladson AR, Shuster WD, Thurston HW, Brown RR (2008) Impediments and solutions to sustainable, watershed- scale urban stormwater management: lessons from Australia and the United States. Environ Manage 42:344–359
36. Brown LD, Khagram S, Moore MH, Frumkin P (2000) Globalization, NGOs, and multi-sectoral relations. In: Nye JS, Donohue JD (eds) Governance in a globalizing world. Brooking Institution Press, Washington, DC, pp 271–296
37. Olsson P, Folke C, Galaz V, Hahn T, Schultz L (2007) Enhancing the fit through adaptive co-management: creating and maintaining bridging functions for matching scales in the Kristianstads Vattenrike Biosphere Reserve Sweden. Ecol Soc 12(1):28. http://www.ecologyandsociety.org/vol12/iss1/art28/
38. Marshall NA (2007) Can policy perception influence social resilience to policy change? Fish Res 86:216–227
39. Lindblom C (1959) The science of muddling through. Public Adm Rev 19:79–88
40. Allison G (1969) Conceptual models and the Cuban missile crisis. Am Political Sci Rev 63:689–718
41. Dietz T, Ostrom E, Stern PC (2003) The struggle to govern the commons. Science 302:1907–1912
42. Hughes TP, Bellwood DR, Folke C, Steneck RS, Wilson J (2005) New paradigms for supporting the resilience of marine ecosystems. Trends Ecol Evol 20:380–386
43. Folke C, Hahn T, Olsson P, Norberg J (2005) Adaptive governance of social- ecological systems. Annu Rev Environ Resour 30:441–473

44. Olsson P, Folke C, Berkes F (2004) Adaptive co-management for building resilience in social-ecological systems. Environ Manage 34:75–90
45. Olsson P, Gunderson LH, Carpenter SR, Ryan P, Lebel L, Folke C, Holling CS (2006) Shooting the rapids: navigating transitions to adaptive governance of social-ecological systems. Ecol Soc 11(1):18. http://www.ecologyandsociety.org/vol11/iss1/art18/
46. Wood BD, Doan A (2003) The politics of problem definition: applying and testing threshold models. Am J Political Sci 47:640–653
47. Gladwell M (2000) The tipping point – how little things can make a big difference. Little, Brown, Boston
48. Zellmer S, Gunderson L (2009) Why resilience may not always be a good thing: lessons in ecosystem restoration from Glen Canyon and the Everglades. Neb Law Rev 87:893–947
49. Kingdon JW (1995) Agendas, alternatives and public policies. Harper Collins College, New York
50. Steelman TA, Tucker DW (2005) The Camino Real: to care for the land and serve the people. In: Brunner RD et al (eds) Adaptive governance: integrating science, policy, and decision making. Columbia University Press, New York, pp 91–130
51. Folke C, Pritchard L, Berkes F, Colding J, Svedin U (2007) The problem of fit between ecosystems and institutions: ten years later. Ecol Soc 12(1):30. http://www.ecologyandsociety.org/vol12/iss1/art30/
52. Ruhl JB (2004) Taking adaptive management seriously: a case study of the Endangered Species Act. Univ Kans Law Rev 52:1249–1284
53. Ruhl JB (1998) The Endangered Species Act and private property: a matter of timing and location. Cornell J Law Public Policy 8:37–54
54. Ruhl JB (2008) Adaptive management for natural resources – inevitable, impossible, or both? Rocky Mountain Law Institute 54:11.01.-06
55. Ruhl JB, Fischman R (2010) Adaptive management in the courts. FSU College of Law, Public Law Research Paper No. 411; Indiana Legal Studies Research Paper No. 154. SSRN: http://ssrn.com/abstract=1542632
56. Karkkainen BC (2005) Panarchy and adaptive change: around the loop and back again. Minn J Law Sci Technol 7:59–77
57. Benson MH (2009) Integrating adaptive management and oil and gas development: existing obstacles and opportunities for reform. Environ Law Rep 39:10962–10978
58. Allen CR, Gunderson LH (2011) Pathology and failure in the design and implementation of adaptive management. J Environ Manag 92:1379–1384
59. Benson MH, Garmestani AS (2011) Embracing panarchy, building resilience and integrating adaptive management through a rebirth of the National Environmental Policy Act. J Environ Manag 92:1420–1427
60. Peterson GD, Cumming GS, Carpenter SR (2003) Scenario planning: a tool for conservation in an uncertain world. Conserv Biol 17:358–366

Chapter 9
Ecotones and Ecological Gradients

Salit Kark

Glossary

Beta-diversity	Also termed species turnover, beta-diversity refers to the change in species as one moves between habitats, communities, or ecosystems.
Divergence-with-gene-flow model of speciation	A model explaining the process of species formation (speciation) in which new species diverge in the face of gene flow; the movement of genes within a group that results from mating with immigrant individuals.
Ecotone	A transition zone between two or more different ecological communities or regions.
Ecotone effect	The pattern of increased species richness (number of species) and abundance in ecotones and the occurrence of unique ecotonal species.
Edge effect	The effect of the juxtaposition of contrasting environments on an ecosystem.
Geographic information systems (GIS)	A computer-based system for creating and managing spatial data and associated attributes. It enables the capture, storage, retrieval, analysis, and display of spatial (location-based) data.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

S. Kark (✉)
The Biodiversity Research Group, Department of Ecology, Evolution, and Behavior,
The Hebrew University of Jerusalem, Jerusalem 91904, Israel

ARC Centre for Environmental Decisions (CEED) School of Biological Sciences,
The University of Queensland, St Lucia, QLD 4072, Australia
e-mail: salit@hebrew.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_9,

Remote sensing	The science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation (e.g., via a satellite image).

Definition of the Subject

Ecotones are areas of transition between ecological communities, ecosystems, or ecological regions (such as Mediterranean and desert). Ecotones often occur along ecological gradients. Such gradients are created as a result of spatial shifts in elevation, climate, soil, and many other environmental factors. Ecotones commonly coincide with areas of sharp climatic transition along environmental gradients. They occur at multiple spatial scales, from continental-scale transitions between major biomes to small-scale ecotones where local vegetation communities and microhabitats coincide. They show a diversity of boundary types that range from natural boundaries (e.g., altitudinal, latitudinal transitions) to human-generated ecotones (e.g., forest clear-cut edges or urban ecotones). Ecotones have been studied in the past four decades in an ecological context and in recent years are receiving increasing attention in the context of biodiversity conservation. Various studies have shown that species richness and abundances tend to peak in ecotonal areas, though exceptions to these patterns occur. Ecotones are "natural laboratories" for studying a range of evolutionary processes, such as the process by which new species form, also termed speciation. Some researchers argue that ecotones deserve high-conservation investment, potentially serving as speciation and biodiversity centers. Because ecotones are often small in size and relatively rich in biodiversity, conservation efforts in these areas may prove to be an efficient and cost-effective conservation strategy.

Introduction

While substantial research in the past decades has focused on distinct ecological regions, communities, and ecosystems, areas of transition between these regions have received much less attention. These areas, nevertheless, sustain rich and unique biodiversity and may have an important role in sustaining future biodiversity. However, studies have shown that species richness and abundances tend to peak in ecotonal areas, though exceptions to this pattern occur. This is likely because ecotonal areas hold species from two or more neighboring communities and possibly due to high spatial heterogeneity and high speciation rates in ecotonal areas. In addition, ecotones have been shown, in some cases, to hold ecotonal species and genotypes that are uniquely characteristic of the ecotone.

Several studies have found evidence that some species are diverging in ecotonal regions, suggesting that transitional areas may serve as speciation centers. Recent work has shown increased genetic and morphological diversity in ecotonal regions compared with neighboring areas. Ecotonal populations often coincide with peripheral populations along a species range. Some researchers suggest that because ecotones often hold marginal populations that depend on other parts of the range for the maintenance of genetic diversity, they do not deserve much conservation interest. Alternatively, in the face of global climate and environmental changes, it has also been argued that ecotones deserve special conservation investment, potentially being speciation and biodiversity hotspots that hold relatively resilient populations. This entry will provide a short overview of the history of ecotone research, recent novel findings in the area and the implications for future conservation, biodiversity, and sustainability research.

Conceptual Framework

Ecotones are areas where ecological communities, ecosystems, or biotic regions coincide. They often occur in areas of steep environmental transition, along environmental gradients. In these transitional regions, the environment rapidly shifts from one type to another based on abiotic (e.g., climatic) and/or biotic (e.g., community structure) factors [13, 19]. The origin of the word "ecotone" is in the Greek roots "oikos" (home) and "tonus" (tension). Many different definitions and terms have been used in the literature to describe ecotones and areas of ecological transition. These include boundary regions, borders, meeting zones, transitional zones, tension zones, zones of intermingling, and zones of transgression [19]. The definition often contrasts ecotones with more homogenous areas found on both sides of the transition or to the landscape as a whole. Ecotones can occur in both terrestrial and aquatic systems, and cover several spatial scales, from large spatial-scale ecotones, where biomes meet [40] to local-scale transitions, such as mountain treelines ([11]; see Fig. 9.1). These areas are sometimes considered to be dynamic zones of interaction between communities, which are unstable over time [19]. As suggested by Odum [30], Ecotones do not simply represent a boundary or an edge; the concept of an ecotone assumes the existence of active interaction between two or more ecosystems with properties that do not exist in either of the adjacent ecosystems. Ecotonal regions show a diversity of transitions that range from natural transitions (e.g., altitudinal and latitudinal transitions) to human-generated ecotones, sometimes termed anthropogenic ecotones (e.g., forest clear-cut edges or urban ecotones) [24], as shown in Fig. 9.1.

History of Research on Ecotones and Biodiversity

For over 100 years, since the late nineteenth century, there has been scientific interest in boundary regions and edges of ecological systems. A conceptual

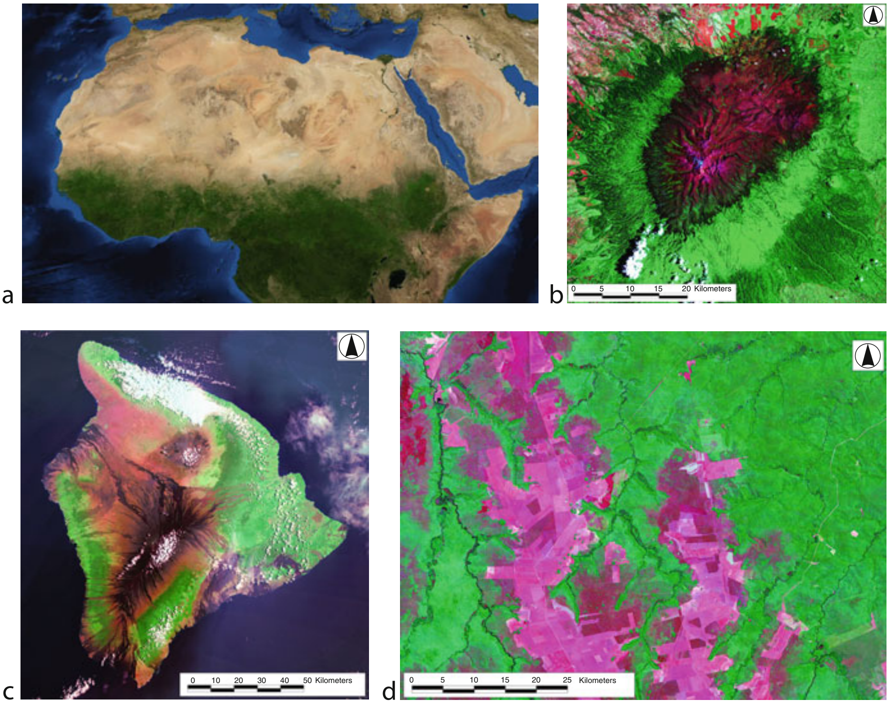


Fig. 9.1 Satellite images showing a variety of natural and human-generated ecotones at several spatial scales: (**a**) Ecotones in the Sahel region of Africa (see text). (**b**)Altitudinal ecotones between vegetation belts in Mt. Kenya. (**c**) Natural and human-made ecotones on Hawaii's big island. Ecotones vary over different slopes and elevations. Note also the sharper ecotones in areas where lava has flowed. (**d**) Human-related ecotones resulting from deforestation (in *pink*) in the Amazon Basin, Brazil. On the far *left* and far *right*, in dark green, natural riparian ecotones can be detected along the rivers. (Source of Fig. **a** Stöckli et al. [39] NASA Blue Marble Next Generation, August 2004. True color composite of MODIS satellite images. Downloaded from: http://visibleearth.nasa.gov/ [42].) (Source of Fig. **b–d** MDA Federal [25] NASA GeoCover circa 1990 Landsat images. False color composite of Landsat bands 7, 4 and 2. *Green shades* represent vegetation, *magenta* and *brown* represent bare soil, *black* represent water bodies, recent lava flows of shadows, *cyan shades* represent snow, and *white* represents clouds. Downloaded from: https://zulu.ssc.nasa.gov/mrsid/ [3].) (All images are public domain. The figures were produced by Noam Levin)

ecological framework for the study of ecotones was given by Odum in 1953. Odum discussed the ecotone effect, which he characterized as increased richness and abundance in ecotones and the occurrence of unique ecotonal species. Until the 1970s, there was considerable interest in ecotones within the scientific community (reviewed by [32]). This interest subsided as focus on more homogenous and well-defined ecosystems and communities (e.g., tropical rain forests and tundra) became common. A revival of research in the field focusing on ecotones and their effects on biodiversity was seen in the late 1980s and 1990s, with the development of new research areas, especially those of landscape ecology and of conservation biology [11, 13]. Studies on ecotones in the 1980s often focused on material flow (e.g., water and nutrients) across communities and on ecosystem processes in these boundary regions [12]. Much of the work focused on wetlands and on riparian zones, where land–water interfaces occur (reviewed in [45]). Later work in the 1990s more directly examined the effect of ecotones on biological diversity, and especially on the relationship between ecotones and processes leading to morphological divergence, patterns of genetic and phenotypic diversity, species richness, rarity, and their conservation implications (reviewed in [17]).

Approaches for Measuring Ecotones

Due to the fact that ecotones can be rarely delimited by a fine line, their measurement and mapping is not simple. A wide range of research approaches and tools have been used to detect and quantify ecotones. These include, among others, simulation modeling, geographic information systems (GIS), remote sensing, and statistical tools that enable quantification and analysis of ecotones of different types and over several spatial scales. Diverse approaches for the quantification of the steepness of gradients exist [17]. Methods for measuring and characterizing ecotones depend on the data available (e.g., quantitative or qualitative, grid- or transect-based data), with one of the simplest approaches, proposed by Womble in 1951, being the quantification of the magnitude of the first and second derivatives (rates of change in a given variable or several variables along a spatial gradient) [44]. These approaches often examine the values of variables in an area (e.g., a 1 × 1 km grid square) relative to its neighboring regions. The basic idea is to detect areas of sharp environmental transition by finding the areas with the highest rate of change in the value of a given variable or several variables between adjacent squares (pixels). Specific software for the detection of boundary regions and analysis is now available (e.g., BoundarySeer: http://www.terraseer.com/products_ boundaryseer.php), enabling more widespread use of advanced statistical tools for the study of areas of transition [5, 8]. These tools can also be applied to the study of ecotonal regions.

In recent years, new approaches to quantify changes in diversity across gradients and boundary regions have been developed and are being applied. Among these is a range of new beta-diversity estimates of species turnover in space [20, 26]. These have been developed in the past decades, since Wilson and Shmida's [43] review on beta-diversity estimates. Beta-diversity and species turnover are often used when studying gradients, and although they do not focus necessarily on ecotonal areas, they can be applied to the study of ecotones.

One of the most promising directions in ecotone and boundary measurement is the use of tools developed in other areas of science. These include fields such as physics, remote sensing, and image analysis, where substantial advancements in boundary detection and gradient quantification methods have been made. Remote-sensing tools, for example, can use data occurring over several orders of magnitude, using satellite-derived data currently available at a resolution of 0.5 m to 100 km. Further application of these tools at multiple spatial resolutions will provide a better understanding of ecotones. Recent work has shown that remote sensing tools can be effectively used to detect ecotones and to predict species richness and rarity (e.g., range size rarity) in ecotones, especially in mountains where latitudinal gradients occur (e.g., [22, 23]).

Patterns of Biodiversity in Ecotones

There have been multiple studies on patterns of biodiversity in ecotonal areas, which have led to a range of results. Recent work is providing increasing evidence that boundary regions between ecological communities can be highly diverse at both the within-species and community levels. Ecotones have been shown to hold especially high biological diversity over several spatial scales, at both the community level (when examining species richness, i.e., the number of species in an area) and at the within-species level (morphological and genetic diversity) as reviewed by Kark and van Rensburg [17]. Other studies, however, have shown conflicting results, making it difficult to generalize without carefully examining each case, community, and region.

Early on, Odum [30] pointed at high species richness and abundance in ecotones and suggested that the ecotonal community commonly contains many species that are characteristic of, and sometimes restricted to, the ecotone. In a continental-scale study of New World birds, Kark et al. [16] examined the relationship between passerine richness and rarity of ~2,300 passeriform species in 4,889 one-degree New World grid cells and the distance of the cells to boundaries between adjacent plant-based ecoregions. They found that areas nearer to transitions between ecoregions had more bird species, and also scored more highly in terms of species rarity. The findings of their work suggest that transitional environments harbor many rare species, in addition to high richness. At the community level, there is also some evidence for high species richness in ecotonal areas in marine systems. For example, van Rensburg et al. [41] showed at the subcontinental scale in South Africa that species richness and range size rarity at a spatial resolution of quarter degree are generally negatively correlated with distance to areas of transition areas between vegetation communities for both birds and frogs. Areas with more range-restricted species were located significantly closer to areas of transition areas between vegetation communities than expected by chance [41]. Similarly, in the Gulf of Aden, Kemp and colleagues [18] found high-reef fish diversity in an ecotone harboring a unique mixing of the three distinct faunas of Oman, the Red Sea, and the Indian Ocean.

Processes Shaping Biodiversity and Sustainability in Ecotones

Ecotones and Evolutionary Processes

Ecotones are “natural laboratories” for studying a range of evolutionary processes, such as the process by which new species form, also termed speciation. This process is of major interest to evolutionary biologists who define three major types of speciation: allopatric, parapatric, and sympatric. These models are based on the degree of geographical subdivision between populations that lead to the formation of new species. Allopatric speciation happens in geographical isolation, and has been for many years considered the major form of speciation. Parapatric speciation occurs in adjacent populations with gene flow among them, often along clines. Sympatric speciation occurs when populations are geographically congruent, and are found in the same area. The study of ecotones has led to a better understanding of the potential importance of parapatric and sympatric speciation as mechanisms for speciation. Ecotones have been proposed by some authors to be centers of evolutionary novelty that maintain evolutionary process, and as regions where parapatric (or sympatric) speciation processes may take place [33]. As such, ecotones and areas of environmental transition have been suggested as natural laboratories where evolutionary processes and barriers to gene flow can be examined [33].

A review by Moritz et al. [27] summarized the major models of evolutionary processes that promote diversification of rain forest faunas. They include the divergence-with-gene-flow model, which suggests that adaptive divergence caused by selection forces occurs across environmental gradients, leading to speciation even in the face of gene flow across ecotones. This means that speciation does not require isolation in cases where selection is strong enough to separate populations. This process is expected to occur especially where very different environments meet in the ecotone, for example, at the border between a dry and wet rain forest. This may mean that even when gene flow continues, strong selection pressures can lead to divergence.

Support for the divergence-with-gene-flow model comes from recent research examining divergence using molecular genetic, phenotypic, and experimental approaches [27, 33, 34, 37]. Smith et al. [37], studying the little greenbul (*Andropadus virens*), a passeriform bird in the rain forest–savannah ecotone region of Cameroon, found especially high morphological divergence in the ecotone. The authors proposed that when the ecotone is large enough, natural selection processes could be strong enough to generate morphological differences similar to those seen in reproductively isolated species even when high rates of gene flow occur. Their data support the divergence-with-gene-flow model of speciation [38], leading them to propose that ecotones may be integral to the production and maintenance of high biodiversity in tropical rain forests. Quantification of morphological and geographic distances in olive sunbird (*Nectarinia olivacea*) populations in West African forests and ecotones revealed similar divergence patterns [38].

Smith studied morphological divergence in another Central African species, the black-bellied seedcracker (*Pyrenestes ostrinus*). Seedcrackers show polymorphism in bill size. A megabilled morph was found in the ecotone that specializes on a very hard-seeded sedge found only in ecotonal areas [27]. This ecotonal megamorph was maintained in the population, despite high levels of gene flow with rain forest populations that had only smaller-billed morphs, owing to its selective advantage for feeding on the hard seeds. Morphological divergence between habitats across an ecotone was also found in leaf-litter skinks (*Carlia rubigularis*) in the wet tropical rain forest of Australia [34]. Adult skinks occurring across sharp ecotones from open (wet sclerophyll) forests to adjacent rain forests showed large morphological and life history differences over short distances despite moderate to high levels of mitochondrial gene flow [34]. Populations occurring across the ecotone had larger size differences than populations located dozens to hundreds of kilometers away that were geographically isolated millions of years ago, suggesting that in this case, and perhaps in others that await research, speciation with gene flow may have importance.

If ecotones regions harbor unique and endemic species and alleles, this may provide support for the notion that these regions may also have importance in speciation. If this is the case, ecotonal regions are expected to contain a preponderance of recently derived species that are yet to expand their ranges (neoendemics). Fjeldså and Rahbek [7] suggested that more recently evolved species are concentrated in transitional ecotones surrounding the main central African rain forest. This reasoning is consistent with the finding that terrestrial ecotones sustain high morphological divergence, providing evidence that current speciation processes may indeed be taking place in these regions [27, 33, 37].

Ecotones, Ecological Gradients, and Species Ranges

One of the important questions that arise is what are the processes that lead to higher species richness in ecotonal regions compared with adjacent areas? Evolutionary processes, as discussed above, may be one explanation. Ecological factors may also have importance in shaping this pattern. Several authors, such as Gosz [11] and Risser [32], have suggested that transitional areas not only share the two types of environments of the habitats that coincide in the ecotone, but also have a unique ecotonal environment. Indeed, Odum [30] proposed that transition zones often support a unique community with characteristics additional to those of the communities that adjoin the ecotone, although also commenting this is by no means a universal phenomenon. Studies testing these predictions show mixed results, some pointing toward the occurrence of ecotonal species, while others not finding evidence for species unique and highly abundant in ecotones. The inconsistency among studies is complicated by the fact that different species, systems, scales, and regions were used in different studies or due to methodological factors, such as differences in sampling and analysis approaches.

In addition, ecotones tend to shift in space and time over several spatial scales [11, 19], as a response to climatic variation, other environmental changes [6, 19, 29], and human activity [10]. Ecotones show high spatial and temporal heterogeneity, which may serve as important factors contributing to their high genetic and species diversity [32]. For example, multiple ecotones can be defined within and around the African Sahel (Fig. 9.1a), depending on the scale of interest and on the definitions used [1]. The different transitions (e.g., that between the Sahel and the desert to the north) experience shifts in time and space, showing high spatiotemporal variability. Another, simpler, process shaping this pattern is that ecotones, comprising meeting areas between adjoining communities, include a combination of species from two or more community types [32]. Ecotonal areas often comprise the edge of the range for species on both sides and are where many peripheral populations occur [17, 35].

An important question is whether populations occurring in ecotonal areas are viable populations that exist over time within ecotones, or are rather present temporarily due to the constant flow of individuals from other parts of their range into the ecotone areas, are not self-sustainable over time, and will disappear if this flow is stopped. Shmida and Wilson [35] proposed that the high number of species in transitional areas could be due to a process they called the mass effect, which is the flow of individuals from favorable to unfavorable areas. For species that reach the margins of their range at the ecotone, this effect may result in some individuals of a given species establishing in ecotonal areas where they cannot maintain viable populations, existing in sinks adjacent to larger source populations [35]. This may lead to increased species richness in ecotonal areas, which is maintained by constant immigration of individuals from more favorable environments. Some evidence for the existence of mass effects can be found in the literature, yet these effects seem to be rather weak, and it is currently unclear whether they can indeed act as a major factor generating high diversity in ecotones [21].

The mass effect, however, does not predict the occurrence of unique or endemic ecotonal species or genotypes. If some species or genotypes are characteristic of an ecotone or occur at the ecotone in higher abundances than in the neighboring habitats, as proposed above, this could suggest that some ecotone populations are ecologically viable. Several recent studies have found peak genetic and morphological diversity within species in ecotone regions, with populations in these regions harboring unique and rare alleles not found elsewhere [17]. For example, a study examining allozyme (protein level) diversity within chukar partridge (*Alectoris chukar*) populations across a rainfall gradient from northern to southern Israel found that the highest levels of diversity occur in the sharp ecotone area between the Mediterranean region and the desert, in the northern Negev area. Populations in this area not only showed higher genetic diversity based on 32 loci (proteins) examined, but also had unique and rare alleles that did not occur elsewhere across the range [14]. The same species also had peak levels of morphological diversity in the ecotone region based on 35 traits and 23 ratios between traits [15]. More studies on genetic and species uniqueness in ecotonal areas are needed to enable generalizations.

Agricultural, Urban and Human-Related Ecotones

The early literature discussing ecotones largely dealt with natural ecotones that are generated by environmental factors such as soils, geology, and climate. More recent research is increasingly including human-related boundaries. Human activity is generating boundaries that did not exist before, changing their steepness and shifting ecotonal location. These boundaries include a diversity of ecotone types, such as forest clear-cut edges, margins between built-up and natural landscapes, and human-generated features, such as lakes and plantations. Research on these human-generated ecotones and their effects on biodiversity are related to a study area that is sometimes termed "countryside biogeography," examining biodiversity in human-dominated landscapes.

Another recently developing research area that is relevant is that of urban ecology. Sharp, human-caused transitions may result from human activities such as urbanization, land-use changes, agriculture, grazing, or burning (see Fig. 9.1). These boundaries occur at multiple spatial scales, ranging from local ecotones between agricultural plots, urban areas, and roads, and their neighboring native habitat, to large-scale ecotones such as shifting desert borders owing to desertification processes and river divergence (Fig. 9.1). These ecotones may be either static and fixed in space or dynamic and shifting in location over space and time.

A substantial amount of work in both natural and human-related landscapes has focused on what has been called the "edge effect." This is the effect of the juxtaposition of contrasting environments on an ecosystem. It refers to how the local environment changes along some type of boundary, or edge and how biodiversity is affected by such edges. This idea is attributed by animal ecologists to Aldo Leopold and his 1933 book *Game Management*. It encompasses a wide range of both biotic and abiotic trends associated with boundaries between adjacent habitat types, natural or anthropogenic. Much of the reference to edge effects in the recent landscape ecology literature has been related to human-caused boundaries, and especially to boundaries between forest fragments and neighboring patches of habitat that have been cleared. Again, there are no clear-cut conclusions as to the effect of human-generated ecotones on biodiversity. The response largely depends on the type of edge and its history as well the species in focus.

Its conservation implications are therefore complex and deserve further scientific attention. A vast amount of research has been done on the effect of forest edges, especially the effect of human-made forest fragments and their edges on biodiversity. A review of edges [28] suggested there are many discrepancies in the literature, and a better understanding and search for general patterns requires a much more mechanistic approach to examine the processes underlying such patterns. In this context, there has been much effort to understand the effect of forest edges on predation, brood parasitism, and the breeding success of birds. Paton [31] found that in the majority of studies, nest success varied near edges with an increase in both depredation and brood parasitism rates. The most conclusive studies suggest that edge effects in birds usually occur within 50 m of an edge [31]. Since these reviews, multiple studies on dozens of species and regions have been conducted.

Implications for Biodiversity, Conservation Planning, and Sustainability

Whereas better understanding of the effects of ecotones on biodiversity is important for evolutionary and ecological pure science purposes, it also has many implications for conservation of biodiversity, sustainability, and practical decision-making. Some scientists have argued that because ecotones hold marginal populations at the edge of the range of many species, where abundances may be lower and populations more prone to local extinctions than other parts of the range, these ecotonal areas have low value for conservation as they will not persist over time, for example, if the surrounding environments become fragmented (e.g., [9]). There has been an interesting discussion in the literature in recent years regarding whether transitional areas are valuable for conservation (e.g., see Smith et al. [36] versus Brooks [4]). Substantial conservation attention has been given in recent years to the understanding and mapping of biodiversity patterns and the underlying processes, and toward predicting the effects of global change. Ecotone and boundary regions, where change, shifts, and variability occur naturally in both space and time, could serve as useful models for understanding, monitoring, and predicting the response of individuals, populations, and communities to changing environments [2, 6, 29]. In addition, while some scientists suggest that ecotone populations are more likely to be negatively affected by climate change [9], other researchers have argued that ecotonal areas sustain populations that are adapted to changing, fluctuating, and unstable environments and, as such, these populations may better be able to persist in the face of predicted change (see further discussion in [17]). This is especially relevant since climate shifts are expected, according to some studies, to be rapid and extreme in boundary regions between ecosystems [2]. By examining changes in ecotone locations over time, these areas may potentially serve as “early warning” indicators of global changes [6, 29]. However, the response depends on the spatial and temporal scales examined and may be a useful indicator mainly at global spatial scales and rather coarse timescales. Therefore, this area deserves further attention owing to the complexity of the factors shaping the location of ecotones in space and time.

Much research in the past decade has focused on prioritizing conservation efforts and determining what areas are most important and valuable for conservation. Two main approaches have been suggested: The first approach includes a search for biodiversity hotspots, or areas with especially high species richness, endemism, and rarity. A second approach has been to select areas that are complementary, and hold biodiversity not present in other areas. Ecotones may provide a unique opportunity to conserve both high species richness and high complementarity. Due to their relatively small size, it may prove a cost-effective strategy to further conserve ecotone regions and to explicitly include them in future systematic conservation planning, given they potentially provide a high return on investment having small area and high biodiversity. Undoubtedly, if these areas have the potential to maintain and to generate species richness as well as unique and novel species and forms, they deserve far greater research attention than they are currently receiving [36]. Nevertheless, conservation plans for ecotones should not be considered independent of their surrounding environments. Ecotonal regions are important for our understanding of evolutionary (such as speciation, divergence with gene flow, and adaptation) and ecological processes shaping biodiversity (e.g., response of populations to fluctuating environments). They may enable us to better predict the responses of populations to environmental change and to further identify previously neglected biodiversity hotspots valuable for biodiversity research and conservation.

Future Directions

Research on the effect of ecotones on biodiversity is an important future direction in the face of global change, including land-use and climate change. Because ecotones are areas where shifts in environmental conditions occur in space and time, they may serve as useful indicators of environmental change and of the response of ecosystems to shifts in climate. The degree of effectiveness of ecotones as early predictors of the impacts of change and the ways that ecological communities and systems respond to change may be an important future direction. Ecotones can sustain unique forms or species that are less abundant or do not occur elsewhere. In addition, studies suggest that ecotones are areas where some populations are diverging to new species in the face of gene flow (across the ecotone). It is important to examine the generality of these patterns and processes in different regions, spatial and temporal scales, and groups.

Another future direction is the application of multi-disciplinary approaches from other fields of science in studying and quantifying ecotones. For example, much work has been done in physics on transitions between states (liquid–solid etc.), which could be applied to work on environmental transitions ecotones. More work on the interactions between native and alien species and the role of ecotones in these interactions should be done to examine the hypothesis that ecotones are barriers to invasion between communities. The importance of ecotones to the generation and conservation of biodiversity, especially in the face of global climate and other environmental changes is an area of potential research and conservation investment that should further be explored and studied. The role of ecotones in promoting sustainability and its emerging science is an open area, which should be further explored in future work.

Bibliography

Primary Literature

1. Agnew CT, Chappell A (2000) Drought in the Sahel. Geo J 48:299–311
2. Allen CD, Breshears DD (1998) Drought-induced shift of a forest-woodland ecotone: rapid landscape response to climate variation. Proc Natl Acad Sci 95:14839–14842
3. Applied Sciences Directorate https://zulu.ssc.nasa.gov/mrsid/. Accessed Nov 2006
4. Brooks TM (2001) Prioritizing hotspots, representing transitions. Trends Ecol Evol 16:673

5. Colwell RK (2006) EstimateS: statistical estimation of species richness and shared species from samples. pp User's guide and application
6. Crumley CL (1993) Analyzing historic ecotonal shifts. Ecol Appl 3:377–384
7. Fjeldså J, Rahbek C (1998) Continent-wide conservation priorities and diversification processes. In: Mace G, Balmford A, Ginsberg J (eds) Conservation in a changing world. Cambridge University Press, Cambridge, pp 107–138
8. Fortin M-J, Olson RJ, Ferson S et al (2000) Issues related to the detection of boundaries. Landsc Ecol 15:453–466
9. Gaston KJ, Rodrigues ASL, van Rensburg BJ, Koleff P, Chown SL (2001) Complementary representation and zones of ecological transition. Ecol Lett 4:4–9
10. Gehrig Fasel J, Guisan A, Zimmermann N (2007) Tree line shifts in the Swiss Alps: climate change or land abandonment? J Veg Sci 18:571–582
11. Gosz JR (1993) Ecotone hierarchies. Ecol Appl 3:369–376
12. Hansen A, Di Castri F (1992) Landscape boundaries: consequences for biotic diversity and ecological flows. Springer, New York
13. Holland M, Risser P, Naiman R (1991) Ecotones. Chapman & Hall, New York
14. Kark S, Alkon PU, Safriel UN, Randi E (1999) Conservation priorities for chukar partridge in Israel based on genetic diversity across an ecological gradient. Conserv Biol 13:542–552
15. Kark S, Mukerji T, Safriel UN, Noy-Meir I, Nissani R, Darvasi A (2002) Peak phenotypic diversity in an ecotone unveiled in the chukar partridge by a novel estimator in a dependent sample (EDS). J Anim Ecol 71:1015–1029
16. Kark S, Allnutt TF, Levin N, Manne LL, Williams PH (2007) The role of transitional areas as avian biodiversity centres. Glob Ecol Biogeogr 16:187–196
17. Kark S, van Rensburg BJ (2006) Ecotones: marginal or central areas of transition? Israel J Ecol Evol 52:29–53
18. Kemp JM (2000) Zoogeography of the coral reef fishes of the north-eastern Gulf of Aden, with eight new records of reef fishes from Arabia. Fauna Arab 18:293–321
19. Kent M, Gill WJ, Weaver RE, Armitage RP (1997) Landscape and plant community boundaries in biogeography. Prog Phys Geogr 21:315–353
20. Koleff P, Gaston KJ, Lennon JJ (2003) Measuring beta diversity for presence–absence data. J Anim Ecol 72:367–382
21. Kunin WE (1998) Biodiversity at the edge: a test of the importance of spatial "mass effects" in the Rothamsted Park Grass experiments. Proc Natl Acad Sci 95:207–212
22. Levanoni O, Levin N, Turbé A, Pe'er G, Kark S (2011) Can we predict butterfly diversity along an elevation gradient from space? Ecography. doi:10.1111/j.1600-0587.2010.06460.x, In press
23. Levin N, Shmida A, Levanoni O, Tamari H, Kark S (2007) Predicting mountain plant richness and rarity from space using satellite derived vegetation indices. Divers Distrib 13:692–703
24. Lloyd KM, McQueen AAM, Lee BJ, Wilson RCB, Walker S, Wilson JB (2000) Evidence on ecotone concepts from switch, environmental and anthropogenic ecotones. J Veg Sci 11:903–910
25. MDA Federal (2004) Landsat Geocover (1990/TM) Edition Mosaics; Tiles N-37-00 and S-37-00 (Mt. Kenya), N-14-15_20 (Hawaii), and S-21-10 (Amazon). US Geological Survey, Sioux Falls, South Dakota
26. Mena J, Vázquez Domínguez E (2005) Species turnover on elevational gradients in small rodents. Glob Ecol Biogeogr 14:539–547
27. Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: an integrated molecular approach. Annu Rev Ecol Syst 31:533–563
28. Murcia C (1995) Edge effects in fragmented forests: implications for conservation. Trends Ecol Evol 10:58–62
29. Neilson RP (1993) Transient ecotone response to climate change: some conceptual and modelling approaches. Ecol Appl 3:385–395
30. Odum EP (1953) Fundamentals of ecology. W. B. Saunders, Philadelphia
31. Paton PWC (1994) The effect of edge on avian nest success: how strong is the evidence? Conserv Biol 8:17–26
32. Risser PG (1995) The status of the science examining ecotones. Bioscience 45:318–325
33. Schilthuizen M (2000) Ecotone: speciation-prone. Trends Ecol Evol 15:130–131

34. Schneider CJ, Smith TB, Larison B, Moritz C et al (1999) A test of alternative models of diversification in tropical rainforests: ecological gradients vs. Rainforest refugia. Proc Natl Acad Sci 96:13869–13873
35. Shmida A, Wilson MV (1985) Biological determinants of species diversity. J Biogeogr 12:1–20
36. Smith TB, Kark S, Schneider CJ, Wayne RK, Moritz C (2001) Biodiversity hotspots and beyond: the need for conserving environmental transitions. Trends Ecol Evol 16:431
37. Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A role for ecotones in generating rainforest biodiversity. Science 276:1855–1857
38. Smith TB, Wayne RK, Girman DJ, Bruford MW (2000) Evaluating the divergence-with-gene-flow model in natural populations: the importance of ecotones in rainforest speciation. In: Bermingham E, Dick CW, Moritz C (eds) Rainforests past and future. University of Chicago Press, Chicago, pp 148–165
39. Stöckli R, Vermote E, Saleous N, Simmon R, Herring D (2005) The blue marble next generation – a true color earth dataset including seasonal dynamics from MODIS. NASA Earth Observatory, Goddard Space Flight Center, Greenbelt
40. van Rensburg BJ, Koleff P, Gaston KJ, Chown SL (2004) Spatial congruence of ecological transition at the regional scale in South Africa. J Biogeogr 31:843–854
41. van Rensburg BJ, Levin N, Kark S (2009) Spatial congruence between ecotones and range-restricted species: implications for conservation biogeography at the sub-continental scale. Divers Distrib 15:379–389
42. Visible Earth http://visibleearth.nasa.gov/
43. Wilson MV, Shmida A (1984) Measuring beta diversity with presence–absence data. J Ecol 72:1055–1064
44. Womble WH (1951) Differential systematics. Science 114:315–322
45. Zalewski M, Schiemer F, Thorpe J (2001) Fish and land-inland water ecotones: overview and synthesis. Ecohydrol Hydrobiol 1:261–266

Books and Reviews

Blake JG, Loiselle BA (2000) Diverstiy of birds along an elevational gradient in the Cordillera Central, Costa Rica. Auk 117:663–686

Kark S (2007, updated 2011) Effects of ecotones on biodiversity. In: Levin S (ed) Encyclopedia of biodiversity. Elsevier, Oxford, p 1–10

Kluge J, Kessler M, Dunn R (2006) What drives elevational patterns of diversity? A test of geometric constraints, climate and species pool effects for pteridophytes on an elevational gradient in Costa Rica. Glob Ecol Biogeogr 15:358–371

Mac Nally R, Molyneux G, Thomson J, Lake P, Read J (2008) Variation in widths of riparian-zone vegetation of higher-elevation streams and implications for conservation management. Plant Ecol 198:89–100

McCain C (2009) Global analysis of bird elevational diversity. Glob Ecol Biogeogr 18:346–360

Odum E, Barrett G (1971) Fundamentals of ecology, 5th edn. W.B. Saunders, Philadelphia

Oommen M, Shanker K (2005) Elevational species richness patterns emerge from multiple local mechanisms in Himalayan woody plants. Ecology 86:3039–3047

Sanchez-Rodriguez J, Baz A (1995) The effects of elevation on the butterfly communities of a Mediterranean mountain, Sierra de Javalambre, central Spain. J Lepid Soc 49:192–207

Senft A (2009) Species diversity patterns at ecotones MSc thesis, University of North Carolina, Chapel Hill

Terborgh J (1971) Distribution on environmetal gradients: theorand a preliminary interpretation of distributional patterns in the avifauna of the Cordillera Vicabamba, Peru. Ecology 52:23–40

Chapter 10
Invasive Species

Anthony Ricciardi

Glossary

Biological invasion	The process by which an organism is introduced to, and establishes a sustainable population in, a region beyond its native range.
Eradication	The managed extirpation of an entire nonnative population.
Impact	The effect of a nonnative species on its environment.
Invasibility	The vulnerability of a habitat, community, or ecosystem to invasion.
Invasion ecology	A multidisciplinary field that examines the causes and consequences of biological invasions.
Invasional meltdown	The phenomenon in which multiple nonnative species facilitate one another's invasion success and impact.
Invasive species	Nonnative species with conspicuously high colonization rates. Such species have the potential to spread over long distances. The term *invasive* is also used (often by policy makers) to describe colonizing species that cause undesirable ecological or economic impacts.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

A. Ricciardi (✉)
Redpath Museum & McGill School of Environment, McGill University,
859 Sherbrooke Street West, Montreal, QC H3A 2K6, Canada
e-mail: tony.ricciardi@mcgill.ca

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_10,

Nonnative species (synonyms: alien, exotic, foreign, nonindigenous)	Species present in a region beyond their historic range.
Propagule pressure	The quantity or rate of nonnative organisms released into an area.

Definition of the Subject

Biological invasion is the process by which a species is introduced, deliberately or inadvertently, into a new geographic region where it proliferates and persists. Outside their historic range (in which they evolved) such species are described as *nonnative* (or nonindigenous, exotic, alien). For a variety of reasons, the vast majority of introduced nonnative organisms fail to persist. Many of those that do establish self-sustaining populations do not spread very far or very fast beyond their point of introduction, and they often do not have conspicuous impacts on their environment. However, a small proportion (but a large and growing number) of nonnative species becomes *invasive* – that is, they may spread aggressively and/or have strong environmental effects. Invasive species are a global problem that threatens native biodiversity, the normal functioning of ecosystems, natural resources, regional economies, and human health. As such, they pose a major concern for conservation and management, and are the focus of a highly productive multidisciplinary field called *invasion ecology*.

Introduction

The potential impact of nonnative species has long been recognized by naturalists. In *The Origin of Species*, Darwin (1859) warned “Let it be remembered how powerful the influence of a single introduced tree or mammal has been shown to be [on native communities].” A century later, Charles Elton’s groundbreaking monograph *The Ecology of Invasions by Animals and Plants* [1] helped inspire two generations of scientists to study what has become one of the world’s most challenging environmental problems.

The major findings of this burgeoning research are summarized in recent texts by Lockwood et al. [2], Davis [3], Blackburn et al. [4], and Richardson [5].

This entry describes the causes and consequences of biological invasions, by synthesizing concepts from population biology, community ecology, evolution, biogeography, and conservation biology. First, the patterns and process of invasion are explored; then, some of its potential ecological and socioeconomic impacts are examined. Some major hypotheses and theoretical concepts explaining patterns of colonization and impact are presented. Next, management approaches to assessing, preventing, and mitigating this problem are considered. The entry ends with a brief glimpse at some of the emerging issues that will likely be the foci of future research.

Pattern and Process in Biological Invasion

The process of invasion comprises a sequence of events involving the transport, introduction, establishment, and spread of organisms into a new region. Organisms in various life stages may be moved by natural dispersal (e.g., passive transport by wind, water currents, or animals; active transport by the organism's own movements) or, far more frequently, by human activities (e.g., transportation systems carrying people or material) across a geographic barrier that previously defined the limits of the historic range of the species. Most organisms will die soon after arrival, or reproduce for only a couple of generations; thus, the vast majority of introduction events fail to produce a sustainable population. If a sufficient number of healthy individuals arrive in a suitable habitat when conditions are favorable, then a self-sustaining population will develop and the species is said to be established. Although populations can sometimes establish from very small numbers, higher numbers of introduced individuals and more frequent introduction events (collectively termed *propagule pressure*) contribute to a higher probability of establishment [6].

In general, the more species introduced to an area, the more that become established in that area [7]. Lonsdale [8] presented an instructive model to describe the number of nonnative species in a region, E:

$$\mathrm{E} = \mathrm{I} \times \mathrm{S}$$

where I is the number of species introduced (*colonization pressure* [7]) and S is the product of the survival rate of each species. S is a function of both the biological traits of the nonnative species and the environmental conditions of the target habitat; for example, all other things being equal, a higher survival rate would result from a closer match between the species' physiological requirements and the prevailing habitat conditions.

There is a variable time lag between initial introduction and establishment, followed by an exponential increase in abundance until the population reaches limits imposed by local abiotic and biotic conditions, at which point population growth diminishes. The range expansion of the species (increase in area occupied per unit time) is correlated with its population growth. The lag phase may range from being negligible (e.g., for a rapidly reproducing species) to extensive – during which the species may remain inconspicuous for years or decades prior to becoming abundant and widespread [9, 10]. For example, the first outbreak of the European gypsy moth (*Lymantria dispar*) in North America occurred two decades after it was initially released. A mussel introduced from the Red Sea remained rare for about 120 years prior to developing dense colonies on the Israeli Mediterranean coast [9].

Recognition of the lag phase phenomenon is critical to management; otherwise, it may lead to inaccurate assessments of benign invasion risk and low impact, as well as missed opportunities to control a nonnative species population while it was still small [10]. Non-mutually exclusive factors contributing to lag phases include: (1) density-dependent (Allee) effects, in which the organism's birth rate is correlated with its population density [11]; (2) adaptation and selection of new genotypes; (3) a change in the composition of the recipient community (e.g., the introduction of a pollinator or seed disperser [12], or the extinction of a dominant resident predator) that triggers the explosive growth of a previously subdued nonnative species; and (4) changing abiotic conditions (e.g., climate change [13]) that release the nonnative species from physiological constraints. Furthermore, the inability to detect an inconspicuous population in its early growth stages is often responsible for a substantial delay in the discovery of a nonnative species. Substantial lags in detection, caused by inadequacies in monitoring and taxonomic expertise, are a major hindrance to effective management [14].

The range expansion of an introduced species tends to fall into a few general patterns, each of which is characterized by an establishment lag phase, an expansion phase, and, when a geographic limit to suitable habitat is realized, a saturation phase [15]. In the simplest pattern, the species expands its range linearly through time; this pattern is the result of random short-distance dispersal outward in all directions through a homogeneous environment, and is often exhibited by rodents such as muskrats. The expanding range is modeled as a circle whose radius increases at a constant rate [16]. The probability of invasion at a given site is inversely proportional to the distance from the edge of the expanding colony and directly proportional to time.

A second pattern is defined by a slow initial rate of linear spread followed by an abrupt shift to a higher linear rate. This biphasic pattern, which has been observed in invasive birds such as the European starling (*Sturnus vulgaris*), occurs when long-distance migrants generate new satellite colonies not far from the primary colony; the coalescence of satellites into the expanding primary colony generates a higher linear rate of expansion. A third pattern occurs when long-distance dispersers create numerous remote satellite colonies that begin to expand their range independent of each other; their continuous coalescence generates an exponential expansion phase, as exhibited by European cheatgrass (*Bromus tectorum*) in North America and tiger pear cactus (*Opuntia aurantiaca*) in South Africa [15, 17]. In this pattern, a prolonged lag phase often occurs prior to conspicuous exponential growth. Genetic adaptation is another mechanism that can produce the enhanced rate of expansion that characterizes the second and third patterns, but the occurrence of long-distance migrants is probably the more common cause. Via long-distance "jumps," migrants may establish satellite colonies that are remote from the expanding edge of the primary colony; the overall rate of range expansion is driven more by the number of these satellite colonies than by their individual size [16]. The pattern is more pronounced where human vectors dominate dispersal, such that there would be multiple introductions of satellite colonies within a region (e.g., the transport of zebra mussels and aquatic weeds between river basins by recreational boats, or introductions of a marine invertebrate along a coastline via ballast water release at various ports). In this case, the probability of dispersal to a given site is nearly independent of time and distance from the primary colony but instead is driven largely by human-mediated dispersal opportunity [18].

Factors Affecting Establishment Success

In addition to propagule pressure, other biotic and abiotic factors have been hypothesized to explain why some species are better invaders, and why some systems are more invaded, than others. Attributes associated with highly invasive species include an ability to rapidly reproduce from small numbers (a high intrinsic rate of population growth), broad environmental tolerance, and mechanisms of exploiting human transportation vectors and human-modified landscapes. A popular view is that generalist species are better invaders than specialists, because the former can thrive in a broader range of habitat conditions (*niche breadth-invasion success hypothesis* [19]). As such, traits that enable species to cope with new environments (e.g., diet breadth, physiological tolerance [20, 21]), or proxy variables that suggest broad tolerance (e.g., latitudinal range [22]), are generally good predictors of invasion success. Among vertebrates, brain size also generally predicts invasion success [23–25], perhaps because it facilitates behavioral flexibility in new environments (but see [26]). Similarly, invasive plants tend to be more phenotypically plastic than noninvasive plants [27]. Traits associated with reproduction are often correlated with the post-establishment success (abundance and range size) of plants [20, 28]. However, the most important factor limiting the large-scale distribution of a species is whether it is valued by humans for domestication [29–32] or, for a species that is not introduced deliberately, whether its life history allows it to be easily transported by human vectors operating on a global scale [33, 34].

Much research on the question of why some communities or systems are more invasible has addressed the concept of *biotic resistance*, which posits that biotic interactions between nonnative species and resident enemies can limit establishment and post-establishment success. The logical extension of this concept is that resident species diversity may act as a barrier to invasion – an idea promoted by Elton [1] to explain the seemingly disproportionate invasibility of species-poor systems such as oceanic islands and highly disturbed areas such as agricultural fields. Most support for Elton's hypothesis is derived from terrestrial plant communities and is equivocal. Over a range of scales, from small garden plots to regional landscapes, positive correlations between native and nonnative species richness have been observed, reflecting shared responses to external variables [35]. Where negative correlations exist, they are found only at local (m^2) scales in experimental manipulations [36]. Numerous studies suggest that competition, herbivory, and native species richness can strongly inhibit the performance (and impact) of nonnative plants following establishment [37, 38], but little evidence suggests that these interactions can prevent establishment when abiotic conditions are favorable and propagule pressure is high. The lesson for managers from these studies is that even highly diverse native communities are often readily invaded by nonnative species, but the reduction of local species richness may accelerate invasion [35].

Most recent studies of invasion mechanisms focus on two popular hypotheses: *fluctuating resource availability* and *enemy release*. The former hypothesis proposes that a system's susceptibility to plant invasions varies with fluctuations in unused resources (e.g., light, water, space, nutrients). Where propagule pressure exists, invasion will be promoted by a sudden increase in resource supply (such as through nutrient pollution) or reduced uptake by resident species (following a disturbance such as clearcutting or fire) [39, 40]. Nutrient-rich habitats do experience more plant invasions, but native plants may not always outperform nonnatives in low-resource conditions [41]. Highly disturbed environments are also believed to be more invasible [1]. Nonnative species may dominate a habitat following a disturbance event that is outside the evolutionary experience of the natives; otherwise, natural disturbance may contribute to a system's resistance to invasion [42].

The enemy release hypothesis attributes the success of nonnative species to their escape from specialized natural enemies upon arrival to a new region, and their inherent advantage over resident competitors that are burdened by their own enemies [43]. One reason why plants that are subject to strong herbivory in their native range can thrive in novel regions is that, in the absence of specialized enemies, they may reallocate the energetic costs of defense toward reproduction and growth, and thus become more competitive [44]. It follows that fast-growing species adapted to resource-rich environments may benefit most from the absence of specialized enemies; thus, multiple mechanisms (enemy release, disturbance, resource addition) may act synergistically to drive such invasions [45].

Modern Invasions as Unprecedented Global Change

The spread of species into regions beyond their native range has accelerated exponentially during the past millennium because of human activities such as agriculture, international travel, and global trade. There is a strong link between trade activity and the global distribution of nonnative species [46, 47]. International trade often involves cargo moved by transoceanic ships, which can carry an enormous number of organisms on their hulls and especially in their ballast tanks. Tens of thousands of ships are estimated to be collectively transporting several thousand species around the planet on any given day [48].

Most countries have recorded the establishment of several hundred nonnative species, including invertebrates, vertebrates, plants, bacteria, and fungi (Fig. 10.1). Human influence is reflected in the improbable composition of modern species assemblages worldwide: African grasses dominate large tracts of the Neotropical region [30], European mammals and birds are abundant in Australia and New Zealand [29, 32], Eurasian invertebrates and fishes dominate food webs in the North American Great Lakes [34], and over 25% of the nonnative species in the Baltic Sea originate from the Pacific and Indian Oceans [50]. Over a decade ago, it was estimated that

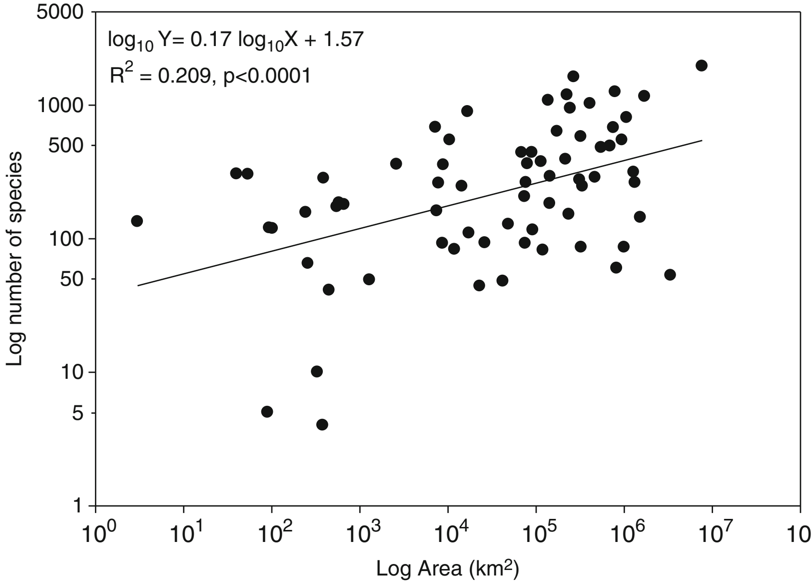


Fig. 10.1 Number of nonnative vascular plant species versus area for regions worldwide (Data from [49]. Line is fitted by least-squares regression)

Table 10.1 Proportion (%) of extant species comprised by established nonnative freshwater fishes, breeding birds, land mammals, and vascular plants in selected regions (Data from [32, 49, 52–57])

Region	Fishes	Birds	Mammals	Plants
Continental areas				
Europe	10	3	19	6
Russia	7	n/a	17	n/a
Southern Africa	11	1	12	4
North America (north of Mexico)	8	4	19	11
South America	<1	<1	4	n/a
Australia	13	6	14	1
Islands				
Puerto Rico	71	35	40	12
Bahamas	14	9	n/a	18
Bermuda	n/a	30	50	65
Hawaii	88	33	89	44
Madagascar	17	2	5	3
Japan	15	2	14	n/a
New Zealand	38	18	40	40

nonnative plants covered at least 3% of the Earth's ice-free land mass, excluding the already immense area under agricultural cultivation [51]. Nonnative species comprise substantial fractions of flora and fauna on continental areas and, especially, on islands (Table 10.1). The majority of these invasions have occurred over the past

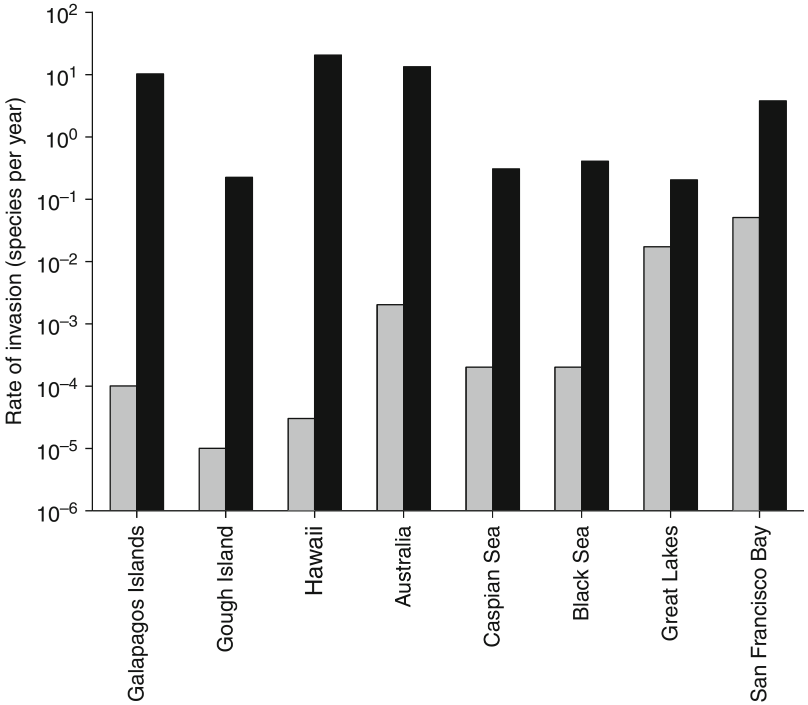


Fig. 10.2 Prehistoric versus modern rates of invasion (number of nonnative species established per year) for various regions. Prehistoric rates (grey bars) are before human settlement and were estimated from the fossil record or by calculating numbers of "native" species (excluding endemics) that have become established in the region over time. Modern rates (black bars) are inferred from discovery rates averaged over the past 40–100 years (Modified from [58])

few centuries, coinciding with steep increases in global trade, human travel, and land use. Invaders are presently colonizing new regions at rates that are several orders of magnitude faster than prior to human arrival (Fig. 10.2). Even the seemingly remote Antarctic continent and its surrounding islands have been colonized by nearly 200 nonnative species of terrestrial plants, invertebrates, and vertebrates within the past two centuries, owing to the effects of scientific exploration, increased accessibility by air and by sea, a burgeoning tourist industry (tens of thousands of visitors annually), and a changing climate [59]. The modern rate and geographic extent of invasion is without historical precedent [58].

Ecological Impacts

Most nonnative species appear to have only minor effects on their invaded systems, but this observation is tempered by two caveats: The impacts of the vast majority of invasions have not been studied [60], and even species that are generally benign can become disruptive at different times or different locations [61]. In many cases, nonnative species can profoundly affect ecosystems by altering community composition, resident species interactions, physical habitat structure, hydrology, nutrient cycling, contaminant cycling, primary production, and natural disturbance (fire, flood, erosion) regimes [17, 62–64]. They can disrupt food webs [65, 66] and plant-animal mutualisms that are crucial for pollination and seed dispersal [67, 68]. Even where environmental stressors such as habitat degradation have already caused population declines of native species, invasions can accelerate these declines [69]. They are a major cause of animal extinctions [70, 71], particularly in insular habitats, such as lakes, river basins, and islands [72, 73]. The invasion-mediated loss of genetically distinct native populations in continental regions has likely been grossly underestimated. There are examples of once widely distributed species being reduced to near extinction as a result of introduced pathogens [17]. Some of the greatest impacts on biodiversity are caused by nonnative predators, and the most conspicuous examples involve introductions to oceanic islands [74, 75] and freshwater ecosystems [76]. Large mammalian herbivores have also had devastating effects on island biodiversity [77, 78]. Other factors contributing to species loss at local to global scales include hybridization [79, 80], competition [69], disease transfer [81], food web alteration [65, 66, 68], and physical habitat alteration [17].

Entire ecosystems may be transformed by invaders that alter resource availability, disturbance regimes, or habitat structure. Some invaders alter the disturbance regime of habitats through fire suppression (e.g., the shrub *Mimosa pigra* in Australian flood plains), fire enhancement (e.g., Eurasian cheatgrass *Bromus tectorum* in the Western United States), increased erosion (e.g., the Australian shrub *Acacia mearnsii* in South Africa), reduced erosion (e.g., exotic plants with extensive root systems that stabilize hills, stream banks, or sand dunes), and increased soil disturbance (e.g., the rooting activities of feral European pigs *Sus scrofa* can destroy the herbaceous understory of a forest, causing soil mineral depletion, rapid organic decomposition, and loss of habitat). Through its filter-feeding activities, the zebra mussel (*Dreissena polymorpha*) has dramatically increased water transparency in North American and European lakes, thus stimulating the growth of benthic algae and macrophytes and altering physical habitat for invertebrates and fishes [82]. In Hawaii, a nitrogen-fixing tree, *Myrica faya*, significantly enriched nutrient-poor volcanic soils at a rate 90-times greater than native plants and thus has a dominant influence on ecosystem properties including soil chemistry and productivity [83]; *Myrica* has also added habitat structure, shading, and high-quality leaf litter that has promoted enhanced populations of nonnative earthworms [84].

Socioeconomic Impacts

The economic value of cultivated nonnative species (such as crop plants) is widely appreciated, but the same cannot be said for the enormous costs incurred by invasions in general. In several countries, nonnative species comprise more than

40% of all harmful weeds, 30% of arthropod pests, and 70% of plant pathogens, and cause substantial losses in total crop production each year [85]. A single invasive forest insect, the emerald ash borer beetle, is projected to cost the United States $10 billion over the next decade [86]. The 2001 outbreak of foot-and-mouth disease in the United Kingdom, linked to illegal meat imports, cost $25 million USD and required the slaughter of ~11 million animals [87]. The annual costs of 16 nonnative species to fisheries, agriculture, and forestry in Canada are projected to be as high as $34 billion CDN [88]. The combined annual costs of biological invasions in the United States, United Kingdom, Australia, India, South Africa, and Brazil are estimated to be $314 billion USD. Assuming similar costs worldwide, the global economic damage attributable to invasions amounts to US $1.4 trillion per year, which constitutes 5% of the global economy [85].

Whereas some nonnative species perform valuable roles, other nonnatives can degrade ecosystem services – including water purification, soil stabilization, agricultural yield, disease regulation, and climate regulation [89]. The conservation of water resources in African countries is threatened by introduced plants [90], whereas pollination services provided by European honeybees are threatened by Asian *Varroa* mites, whose parasitism has destroyed entire hives [91]. Animal (including human) health, in general, is threatened by invasions that spread parasites, diseases, and their vectors (e.g., mosquitoes [92]). Invasions can also alter the transmission of parasites to humans by introducing hosts to novel regions [93]. About 100 species (~6%) of nonnative invertebrates (e.g., spiders, mosquitoes, nematodes) in Europe adversely affect human or animal health, and these are a subset of ~1,300 nonnative species in the region that have documented socioeconomic impacts [94]. Climate change is expected to drive a new wave of such invasions, as suggested by the recent occurrence in Northern Europe of the tropical virus that causes "bluetongue disease" that resulted from the introduction of infected livestock from a Mediterranean country [95].

Management of Invasions

Risk Assessment

Managers have few tools for prioritizing invasion threats because reliable predictive methods are scarce (but see [96, 97]). Progress in developing a predictive understanding of impact has been hampered by the lack of standardized metrics. Parker et al. [60] proposed a metric for impact (I) that can be compared across species and invaded sites:

$$\mathrm{I} = \mathrm{R} \times \mathrm{A} \times \mathrm{E}$$

where R is the total area occupied by the nonnative species in its invaded range, A is its abundance (in numbers or biomass per square meters) in the invaded range, and E is its per-capita effect based on the functional ecology and behavior of individuals (e.g., filtration rate of mussels, functional response of predators, rate of habitat conversion for ecosystem engineers). Data on per-capita effects are often scarce, but inferences regarding the magnitude of impact may be drawn from abundance, which has been shown to be a useful predictor of impact [61]. Range size, in contrast, may not necessarily be a good predictor. Beyond the trivial expectation that the impacts of an invading species accumulate as it occupies more territory, there is no statistical correlation between the invasion success of a species (i.e., its rate of establishment success or spread) and the magnitude of its impact [98]. Even relatively poor invaders can have strong local impacts on native populations (e.g., the Asian clam *Potamocorbula amurensis*; Atlantic salmon *Salmo salar*), whereas highly successful colonizers do not necessarily displace native species (e.g., freshwater jellyfish *Craspedacusta sowerbyi*). One generalization that has emerged from numerous case studies is that high-impact invaders often represent novel life forms in the invaded system. They acquire and use resources differently than resident species, possess defense mechanisms and "weapons" that are foreign to the invaded community [99], and may have predatory capabilities to which residents are poorly adapted. Such species tend to belong to taxonomic or functional groups that were not present in the ecosystem prior to invasion [100–102]. As such, the phylogenetic distinctiveness of the invader in its novel environment might be an indicator of its impact potential [101, 102].

A major challenge to prediction is context-dependent variation generated by site-specific environmental factors [60, 61]. The best predictor of the colonization success and impact of an introduced plant or animal is its invasion history [20, 61]. Although impacts vary across a heterogeneous environment, models may be developed to predict the impact (or abundance) of a species with a well-documented impact history [61], but the predictive power of such models is diminished at sites that have been highly invaded. Nonnative species can interact in multiple ways to produce unpredictable effects [12, 75], sometimes by facilitating each other's spread and impact (i.e., *invasional meltdown* [103]).

Prevention

Given the growing frequency of invasions, their profound impacts, and the substantive resources required to control rapidly spreading species after they become established, the most cost-effective management strategy is prevention [14]. Arguably, invasions warrant similar investments in preparedness and response planning as natural disasters; despite being slower in their onset, invasions have more persistent impacts and a greater scope of ecological and economic damage than natural disasters [104].

Prevention involves controlling either species entry or establishment. Preventing entry of nonnative species begins with the identification and control of dominant transportation vectors and pathways [14]. The effectiveness of vector-control policies requires rigorous inspection, enforcement, evaluation, and – where necessary – refinement, as has been demonstrated by the evolution of a management program to control ballast water–mediated invasions in the Great Lakes [105]. An additional preventative approach is to manage ecosystems so as to reduce their vulnerability to invasions – e.g., via restoration of intact native communities in degraded areas, managed disturbance (e.g., fire, river flow) regimes, and manipulation of resource supply (nutrients, water supply) [14, 106]. Cultivated systems can be designed with resistance in mind; for example, the use of polycultures (e.g., diversified crops, mixed forest stands) has been demonstrated to reduce harmful outbreaks of invasive pests [107]. The spatial modification of habitats (such as the use of small-scale dispersal barriers) may also be employed to limit colonization [11].

Eradication

The Convention on Biological Diversity [article 8(h)] directs signatory nations to "prevent the introduction of, control or eradicate those alien species which threaten ecosystems." Eradication, the removal of a nonnative population, can lead to the recovery of previously threatened native species [108, 109]. Several conditions must be met for an eradication program to be successful [110]: (1) The target species must be detected at low densities. (2) Its biology must make it susceptible to control measures. (3) Resources must be sufficient to complete the project. (4) Managers must have the authority and public support to take all necessary steps. (5) Re-invasion must be prevented. Also influencing the success of eradication are the reproductive and dispersal capabilities of the invader, both of which determine how fast it will spread. The probability of success is highest in the initial stages of invasion when spatial spread is still limited; hence, early detection and rapid response are crucial, particularly for species that can reproduce and disperse rapidly [14].

Owing to the indirect effects of nonnative species, eradication can have unanticipated negative consequences. Where multiple invaders exist, particularly in simple food webs (e.g., on islands), the removal of a nonnative predator or herbivore can cause the proliferation of a second invader that was previously controlled by the target species through top-down regulation [111, 112]. For example, the eradication of feral cats from Macquarie Island led to a population explosion of an invasive herbivore – European rabbit [112]. The explosion of rabbits was accompanied by large-scale habitat alteration characterized by a shift in vegetation that favored fast-growing plants, some of which themselves were nonnative. Similarly, the removal of cats from Little Barrier Island, New Zealand, released the introduced Pacific rat (*Rattus exulans*) from top-down control and led to a reduction in the breeding success of an endangered endemic seabird (Cook's petrel, *Pterodroma cookii*), apparently due to nest predation by the rat; subsequent eradication of the rat was followed by a rapid rise in the seabird's breeding success [111]. Additional effects of eradication on multiply invaded systems might be to increase predation pressure on natives as a result of nonnative predators shifting their diets following the removal of nonnative prey, or to release one or more nonnative species from competition by removing a superior competitor.

Maintenance Control

When dealing with nonnative species with strong Allee effects, eradication may involve culling individuals to bring a population below sustainable levels [11]. If eradication fails, or is impossible, the next option is maintenance control of the invader at acceptable population levels, using mechanical, chemical, or biological control methods. Mechanical control, such as hunting, may be particularly effective on islands and other geographically restricted areas. Chemical control involves the application of pesticides to reduce the abundance of a target species, but high economic costs and human health risks constrain the application of chemicals over large areas. Moreover, pesticides often impact nontarget species (including native competitors), sometimes to the benefit of the target itself [113].

Biological control involves the introduction of a nonnative species (usually a predator, herbivore, or parasite) to reduce an established nonnative pest to less harmful densities. This technology is considered to be a more desirable alternative to pesticide use, despite its potential for unanticipated consequences. Because the introduced agents can disperse beyond the target area and evolve to exploit new hosts, nontarget species may be attacked and even driven to extinction [17, 114]. The assumption underlying biological control is that nonnative species proliferate to harmful levels because they have escaped their natural enemies. However, indirect (e.g., competitive) effects may sometimes be more important than top-down consumer regulation. Under these situations, the introduction of a biological control species may have a counterproductive effect [115]. Difficulties in predicting such complex community interactions can obviously compromise ecological risk assessments.

Future Directions

The questions underlying invasion ecology – that is, why some species are more successful and have greater impact than others, why some systems are more vulnerable to invasion, and how ecosystem functions and services are affected by invasion – are clearly of societal importance and will remain relevant in the future, as invasive species are increasingly viewed as a biosecurity issue [87]. The extent and impact of invasions will be further exacerbated by climate change, and synergies between nonnative species and other human-mediated stressors will become more frequent. Future research foci will include the consequences associated with cultivation of novel biofuels and bioenergy crops [116] and the expanded use of genetically modified organisms [117]. Moreover, there may be increasing interest among conservation biologists to relocate native species deemed to be threatened by climate change or other stressors, and some plants and animals could be moved well beyond their historical ranges [73]. Each of these practices will have potentially high ecological risks whose assessment will require more powerful forecasting methods than are currently available. Thus, we can anticipate a growing need for invasion ecology to develop a more predictive understanding of the impact of nonnative organisms.

Bibliography

Primary Literature

1. Elton CS (1958) The ecology of invasions by animals and plants. Methuen, London
2. Lockwood JL, Hoopes MF, Marchetti MP (2007) Invasion ecology. Blackwell, Oxford
3. Davis MA (2009) Invasion biology. Oxford University Press, Oxford
4. Blackburn TM, Lockwood JL, Cassey P (2009) Avian invasions. Oxford University Press, Oxford
5. Richardson DM (ed) (2011) Fifty years of invasion ecology – the legacy of Charles Elton. Wiley-Blackwell, Chichester
6. Simberloff D (2009) The role of propagule pressure in biological invasions. Annu Rev Ecol Evol System 40:81–102
7. Lockwood JL, Cassey P, Blackburn TM (2009) The more you introduce the more you get: the role of propagule and colonization pressure in invasion success. Divers Distrib 15:904–910
8. Lonsdale WM (1999) Global patterns of plant invasions and the concept of invasibility. Ecology 80:1522–1536
9. Rilov G, Benayahu Y, Gasith A (2004) Prolonged lag in population outbreak of an invasive mussel: a shifting habitat model. Biol Invasions 6:347–364
10. Crooks JA (2005) Lag times and exotic species: the ecology and management of biological invasions in slow-motion. Ecoscience 12:316–329
11. Tobin PC, Berec L, Liebhold AM (2011) Exploiting allee effects for managing biological invasions. Ecol Lett 14:615–624
12. Richardson DM, Allsopp N, D'Antonio CM, Milton SJ, Rejmànek M (2000) Plant invasions – the role of mutualisms. Biol Rev 75:65–93
13. Witte S, Buschbaum C, van Beusekom EE, Reise K (2010) Does climatic warming explain why an introduced barnacle finally takes over after a lag of more than 50 years? Biol Invasions 12:3579–3589
14. Hulme PE (2006) Beyond control: wider implications for the management of biological invasions. J Appl Ecol 43:835–847
15. Shigesada N, Kawasaki K (1997) Biological invasions: theory and practice. Oxford University Press, Oxford
16. Hengeveld R (1989) Dynamics of biological invasions. Chapman and Hall, New York
17. Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzazz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecol Appl 10:689–710
18. MacIsaac HJ, Grigorovich IA, Ricciardi A (2001) Reassessment of species invasions concepts: the Great Lakes basin as an example. Biol Invasions 3:405–416
19. Vasquez DP (2006) Exploring the relationship between niche breadth and invasion success. In: Cadotte M, McMahon SM, Fukami T (eds) Conceptual ecology and invasion ecology. Springer, Dordrecht, pp 307–322
20. Hayes KR, Barry SC (2008) Are there any consistent predictors of invasion success? Biol Invasions 10:483–506
21. Blackburn TM, Cassey P, Lockwood JL (2009) The role of species traits in the establishment success of exotic birds. Glob Chang Biol 15:2852–2860
22. Goodwin BJ, McAllister AJ, Fahrig L (1999) Predicting invasiveness of plant species based on biological information. Conserv Biol 13:422–426

23. Sol D, Timmermans S, Lefebvre L (2002) Behavioural flexibility and invasion success in birds. Anim Behav 63:495–502
24. Sol D, Bacher S, Reader SM, Lefebvre L (2008) Brain size predicts the success of mammal species introduced to novel environments. Am Nat 172:S63–S71
25. Amiel JJ, Ringley R, Shine R (2011) Smart moves: effects of relative brain size on establishment success of invasive amphibians and reptiles. PLoS One 6(4):e18277
26. Drake JM (2007) Parental investment and fecundity, but not brain size, are associated with establishment success in introduced fishes. Funct Ecol 21:963–968
27. Davidson AM, Jennions M, Nicotra AB (2011) Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. Ecol Lett 14:419–431
28. Ordonez A, Wright IJ, Olff H (2010) Functional differences between native and alien species: a global-scale comparison. Funct Ecol 24:1353–1361
29. Long JL (1981) Introduced birds of the world. Universe Books, New York
30. Mack RN, Lonsdale WM (2001) Humans as global plant dispersers: getting more than we bargained for. BioScience 51:95–102
31. Reichard SH, White P (2001) Horticulture as a pathway of invasive plant introductions in the United States. BioScience 51:103–113
32. Long JL (2003) Introduced mammals of the world. CSIRO Publishers, Collingwood
33. Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: insights from Argentine ants. Proc Natl Acad Sci USA 98:1095–1100
34. Ricciardi A (2006) Patterns of invasion in the Laurentian Great Lakes in relation to changes in vector activity. Divers Distrib 12:425–433
35. Fridley JD, Stachowicz JJ, Naeem S, Sax DF, Seabloom EW, Smith MD, Stohlgren TJ, Tilman D, Von Holle B (2007) The invasion paradox: reconciling pattern and process in species invasions. Ecology 88:3–17
36. Herben T, Mandák B, Bímova K, Münzbergova Z (2004) Invasibility and species richness of a community: a neutral model and a survey of published data. Ecology 85:3223–3233
37. Levine JM, Adler PB, Yelenik SG (2004) A meta-analysis of biotic resistance to exotic plant invasions. Ecol Lett 7:975–989
38. Olofsson J, Oksanen L, Callaghan T, Hulme PE, Oksanen T, Suominen O (2009) Herbivores inhibit climate-driven shrub expansion on the tundra. Glob Chang Biol 15:2681–2693
39. Davis MA, Grime JP, Thompson K (2000) Fluctuating resources in plant communities: a general theory of invasibility. J Ecol 88:528–534
40. Davis MA, Pelsor M (2001) Experimental support for a resource-based mechanistic model of invasibility. Ecol Lett 4:421–428
41. Funk JL, Vitousek PM (2007) Resource-use efficiency and plant invasion in low-resource systems. Nature 446:1079–1081
42. Leprieur F, Hickey MA, Arbuckle CJ, Closs GP, Brosse A, Townsend CR (2006) Hydrological disturbance benefits a native fish at the expense of an exotic fish. J Appl Ecol 43:930–939
43. Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. Trends Ecol Evol 17:164–170
44. Blossey B, Nötzold R (1995) Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. J Ecol 83:887–889
45. Blumenthal D, Mitchell CE, Pyšek P, Jarošik V (2009) Synergy between pathogen release and resource availability in plant invasion. Proc Natl Acad Sci USA 106:7899–7904
46. Westphal MI, Browne M, MacKinnon K, Noble I (2008) The link between international trade and the global distribution of invasive alien species. Biol Invasions 10:391–398
47. Lin W, Cheng X, Xu R (2011) Impact of different economic factors on biological invasions on the global scale. PLoS One 6(4):e18797
48. Carlton JT (1999) The scale and ecological consequences of biological invasion in the World's oceans. In: Sandlund OT, Schei PJ, Viken Å (eds) Invasive species and biodiversity management. Kluwer, Dordrecht, pp 195–212
49. Vitousek PM, D'Antonio CM, Loope LL, Rejmanek M, Westbrooks R (1997) Introduced species: a significant component of human-caused global change. N Z J Ecol 21:1–16

50. Leppäkoski E, Olenin S (2000) Non-native species and rates of spread: lessons from the brackish Baltic Sea. Biol Invasions 2:151–163
51. Mack RN (1997) Plant invasions: early and continuing expressions of global change. In: Huntley B, Cramer W, Morgan AV, Prentice HC, Allen JRM (eds) Past and future rapid environmental changes: the spatial and evolutionary responses of terrestrial biota. Springer, Berlin, pp 205–216
52. Bogustkaya NG, Naseka AM (2002) Freshwater fishes of Russia database. http://www.zin.ru/animalia/pisces/. Accessed 22 July 2011
53. Callmander MW, Phillipson PB, Schatz GE, Andriambololonera S, Rabarimanarivo M, Rakotonirina N, Raharimampionona J, Chatelain C, Gautier L, Lowry PP (2011) The endemic and non-endemic vascular flora of Madagascar updated. Plant Ecol Evol 144:121–125
54. Froese R, Pauly D (2011) FishBase. www.fishbase.org, version (06/2011). Accessed 11 July 2011
55. Hockey PAR, Dean WRJ, Ryan PG (2005) Roberts – birds of southern Africa, 7th edn. The Trustees of the John Voelcker Bird book Fund, Cape Town
56. Koehn JD, MacKenzie RF (2004) Priority management actions for alien freshwater fish species in Australia. N Z J Mar Freshw Res 38:457–472
57. Lehtonen H (2002) Alien freshwater fishes of Europe. In: Leppakoski E, Gollasch S, Olenin S (eds) Invasive aquatic species of Europe: distribution, impacts and management. Kluwer, Dordrecht
58. Ricciardi A (2007) Are modern biological invasions an unprecedented form of global change? Conserv Biol 21:329–336
59. Frenot Y, Chown SL, Whinam J, Selkirk PM, Convey P, Skotnicki M, Bergstrom DM (2005) Biological invasions in the Antarctic: extent, impacts and implications. Biol Rev 80:45–72
60. Parker IM, Simberloff D, Lonsdale WM, Goodell K, Wonham M, Kareiva PM, Williamson MH, Von Holle B, Moyle PB, Byers JE, Goldwasser L (1999) Impact: toward a framework for understanding the ecological effects of invaders. Biol Invasions 1:3–19
61. Ricciardi A (2003) Predicting the impacts of an introduced species from its invasion history: an empirical approach applied to zebra mussel invasions. Freshw Biol 48:972–981
62. Brooks ML, D'Antonio CM, Richardson DM, Grace JB, Keeley JE, DiTomaso JM, Hobbs RJ, Pellant M, Pyke D (2004) Effects of invasive alien plants on fire regimes. BioScience 54:677–688
63. Croll DA, Maron JL, Estes JA, Danner EM, Byrd GV (2005) Introduced predators transform subarctic islands from grassland to tundra. Science 307:1959–1961
64. Simberloff D (2011) How common are invasion-induced ecosystem impacts? Biol Invasions 13:1255–1268
65. Spencer CN, McClelland BR, Stanford JA (1991) Shrimp stocking, salmon collapse, and eagle displacement. BioScience 41:14–21
66. Burghardt KT, Tallamy DW, Philips C, Shropshire KJ (2010) Non-native plants reduce abundance, richness, and host specialization in lepidopteran communities. Ecosphere 1(5):1–22
67. Traveset A, Richardson DM (2006) Biological invasions as disruptors of plant reproductive mutualisms. Trends Ecol Evol 21:208–216
68. Sekercioglu CH (2011) Functional extinctions of bird pollinators cause plant declines. Science 331:1019–1020
69. Ricciardi A (2004) Assessing species invasions as a cause of extinction. Trends Ecol Evol 19:619
70. Clavero M, García-Berthou E (2005) Invasive species are a leading cause of animal extinctions. Trends Ecol Evol 20:110
71. Clavero M, Brotons L, Pons P, Sol D (2009) Prominent role of invasive species in avian biodiversity loss. Biol Conserv 142:2043–2049
72. Ebenhard T (1988) Introduced birds and mammals. Swed Wildl Res (Viltrevy) 13(4):1–107
73. Ricciardi A, Simberloff D (2009) Assisted colonization is not a viable conservation strategy. Trends Ecol Evol 24:248–253

74. Fritts TH, Rodda GH (1998) The role of introduced species in the degradation of island ecosystems: a case history of Guam. Annu Rev Ecol Syst 29:113–140
75. Blackburn TM, Petchey OL, Cassey P, Gaston KJ (2005) Functional diversity of mammalian predators and extinction in island birds. Ecology 86:2916–2923
76. Witte F, Goldschmidt T, Wanink J, Vanoijen M, Goudswaard K, Wittemaas E, Bouton N (1992) The destruction of an endemic species flock – quantitative data on the decline of the haplochromine cichlids of Lake Victoria. Environ Biol Fish 34:1–28
77. Spear D, Chown SL (2009) Non-indigenous ungulates as a threat to biodiversity. J Zool 279:1–17
78. Carrete M, Serrano D, Illera JC, López G, Vögeli M, Delgado A, Tella JL (2009) Goats, birds, and emergent diseases: apparent and hidden effects of exotic species in an island environment. Ecol Appl 19:840–853
79. Simberloff D (2006) Hybridization between native and introduced wildlife species: importance for conservation. Wildl Biol 2:143–150
80. Ayres DR, Zaremba K, Strong DR (2004) Extinction of a common native species by hybridization with an invasive congener. Weed Technol 18S:1288–1291
81. Wyatt KB, Campos PF, Gilbert MTP, Kolokotronis S-O, Hynes WH, DeSalle R, Ball SJ, Daszak P, MacPhee RDE, Greenwood AD (2008) Historical mammal extinction on Christmas Island (Indian Ocean) correlates with introduced infectious disease. PLoS One 3(11): e3602
82. Higgins SN, Vander Zanden MJ (2010) What a difference a species makes: a meta–analysis of dreissenid mussel impacts on freshwater ecosystems. Ecol Monogr 80:179–196
83. Vitousek PM, Whiteaker LR, Mueller-Dombois D, Matson PA (1987) Biological invasion by *Myrica faya* alters ecosystem development in Hawai'i. Science 238:802–804
84. Aplet GH (1990) Alteration of earthworm community biomass by the alien *Myrica faya* in Hawaii. Oecologia 82:414–416
85. Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, Wong E, Russel L, Zern J, Aquino T, Tsomondo T (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. Agric Ecosyst Environ 84:1–20
86. Kovacs KF, Haight RG, McCullough DG, Mercader RJ, Siegert NW, Liebhold AM (2010) Cost of potential emerald ash borer damage in U.S. communities, 2009–2019. Ecol Econ 69:569–578
87. Chomel BB, Sun B (2010) Bioterrorism and invasive species. Rev Sci Tech 29:193–199
88. Colautti RI, Bailey SA, van Overdijk CDA, Amundsen K, MacIsaac HJ (2006) Characterised and projected costs of nonindigenous species in Canada. Biol Invasions 8:45–59
89. Pejchar L, Mooney HA (2009) Invasive species, ecosystem services and human well-being. Trends Ecol Evol 24:497–504
90. Le Maitre DC, Versfeld DB, Chapman RA (2000) The impact of invading alien plants on surface water resources in South Africa: a preliminary assessment. Water SA 26:397–408
91. Cook DC, Thomas MB, Cunningham SA, Anderson DL, De Barro PJ (2007) Predicting the economic impact of an invasive species on an ecosystem service. Ecol Appl 17:1832–1840
92. Lounibos LP (2002) Invasions by insect vectors of human disease. Annu Rev Entomol 47:233–266
93. Lv S, Zhang Y, Steinmann P, Yang G-J, Yang K, Zhou X-N, Utzinger J (2011) The emergence of angiostrongyliasis in the People's Republic of China: the interplay between invasive snails, climate change and transmission dynamics. Freshw Biol 56:717–734
94. Vilà M, Basnou C, Pyšek P, Josefsson M, Genovesi P, Gollasch S, Nentwig W, Olenin S, Roques A, Roy D, Hulme PE (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. Front Ecol Environ 8:135–144
95. Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PPC, Baylis M (2005) Climate change and the recent emergence of bluetongue in Europe. Nat Rev Microbiol 3:171–181
96. Reichard S, Hamilton CW (1997) Predicting invasions of woody plants introduced into North America. Conserv Biol 11:193–203
97. Kolar CS, Lodge DM (2002) Ecological predictions and risk assessments for alien species. Science 298:1233–1236

98. Ricciardi A, Cohen J (2007) The invasiveness of an introduced species does not predict its impact. Biol Invasions 9:309–315
99. Callaway RM, Ridenour WM (2004) Novel weapons: invasive success and the evolution of increased competitive ability. Front Ecol Environ 2:436–443
100. Short J, Kinnear JE, Robley A (2002) Surplus killing by introduced predators in Australia – evidence for ineffective anti-predator adaptations in native prey species? Biol Conserv 103:283–301
101. Ricciardi A, Atkinson SK (2004) Distinctiveness magnifies the impact of biological invaders in aquatic ecosystems. Ecol Lett 7:781–784
102. Strauss SY, Webb CO, Salamin N (2007) Exotic taxa less related to native species are more invasive. Proc Natl Acad Sci U S A 103:5841–5845
103. Simberloff D, Von Holle B (1999) Positive interactions of nonindigenous species: invasional meltdown? Biol Invasions 1:21–32
104. Ricciardi A, Palmer ME, Yan ND (2011) Should biological invasions be managed as natural disasters? BioScience 61:312–317
105. Bailey SA, Deneau MG, Jean L, Wiley CJ, Leung B, MacIsaac HJ (2001) Evaluating efficacy of an environmental policy to prevent biological invasions. Environ Sci Technol 45:2554–2561
106. Huston MA (2004) Management strategies for plant invasions: manipulating productivity, disturbance, and competition. Divers Distrib 10:167–178
107. Risch SJ, Andow D, Altieri MA (1983) Agroecosystem diversity and pest control: data, tentative conclusions, and new directions. Environ Entomol 12:625–629
108. Pascal M, Siorat F, Lorvelec O, Yésou P, Simberloff D (2005) A pleasing consequence of Norway rat eradication: two shrew species recover. Divers Distrib 11:193–198
109. Donlan CJ, Campbell K, Cabrera W, Lavoie C, Carrion V, Cruz F (2007) Recovery of the Galapagos rail (*Laterallus spilonotus*) following the removal of invasive mammals. Biol Conserv 138:520–524
110. Myers JH, Simberloff D, Kuris AM, Carey JR (2000) Eradication revisited: dealing with exotic species. Trends Ecol Evol 15:316–320
111. Rayner MJ, Hauber ME, Imber MJ, Stamp RK, Clout MN (2007) Spatial heterogeneity of mesopredator release within an oceanic island system. Proc Natl Acad Sci U S A 104:20862–20865
112. Bergstrom DM, Lucieer A, Kiefer K, Wasley J, Belbin L, Pedersen TK, Chown SL (2009) Indirect effects of invasive species removal devastate World Heritage Island. J Appl Ecol 46:73–81
113. Rinella MJ, Maxwell BD, Fay PK, Weaver T, Sheley RL (2009) Control effort exacerbates invasive-species problem. Ecol Appl 19:155–162
114. Louda SM, Arnett AE, Rand TA, Russell FL (2003) Invasiveness of some biological control insects and adequacy of their ecological risk assessment and regulation. Conserv Biol 17:73–82
115. Callaway RM, DeLuca TH, Belliveau WM (1999) Biological-control herbivores may increase competitive ability if the noxious weed *Centaura maculosa*. Ecology 80:1196–1201
116. Lonsdale WM, FitzGibbon F (2011) The known unknowns – managing the invasion risk from biofuels. Curr Opin Environ Sustain 3:31–35
117. Andow DA, Zwahlen C (2006) Assessing environmental risks of transgenic plants. Ecol Lett 9:196–214

Books and Reviews

Catford JA, Jansson R, Nilsson C (2009) Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. Divers Distrib 15:22–40

Chapter 11
Landscape Ecology

Jianguo (Jingle) Wu

Glossary

Landscape	A geographic area in which variables of interest are spatially heterogeneous. The boundary of a landscape may be delineated based on geographic, ecological, or administrative units (e.g., a watershed, an urban area, or a county) which are relevant to the research questions and objectives.
Landscape connectivity	The degree of a landscape to facilitate or impede the exchange of organisms, energy, material, and information among landscape elements. This is sometimes referred to as landscape functional connectivity, which is a function of both landscape structural connectivity and the movement characteristics of the species or process under consideration. Landscape structural connectivity is simply a measure of how spatially connected the elements in a landscape are, without reference to any particular ecological process.
Landscape ecology	The science of studying and improving the relationship between spatial pattern and ecological processes in a landscape on multiple scales. Landscape ecology studies the structure, function, and dynamics of landscapes of different kinds, including natural, seminatural, agricultural, and urban landscapes.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

J. (Jingle) Wu (✉)
School of Life Sciences and Global Institute of Sustainability, Arizona State University,
P.O. Box 874501, 85287-4501 Tempe, AZ, USA
e-mail: Jingle.Wu@asu.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_11,

Landscape fragmentation	The breaking-up of landscape into smaller patches by anthropogenic and natural forces or the introduction of barriers that impede exchange of organisms, energy, material, and information across a landscape. Habitat fragmentation is a similar term to landscape fragmentation, but has a more explicit focus on changes in habitat relevant for organisms of interest.
Landscape pattern	The composition (diversity and relative abundances) and configuration (shape, size, and spatial arrangement) of landscape elements, including both spatial patchiness and gradients.
Landscape function	The horizontal and vertical exchanges of organisms, energy, material, and information in a landscape.
Landscape structure	The composition and spatial arrangement of landscape elements – including patches, corridors, and the matrix.
Landscape dynamics	Temporal changes in the structure and function of a landscape, driven by natural and anthropogenic processes.
Landscape sustainability	The ability of a landscape to maintain its basic environmental, economic, and social functions under ever-changing conditions driven by human activities and environmental changes. Landscape sustainability emphasizes the optimization of the composition and spatial configuration of the landscape so as to achieve a high level of resilience or persistency.
Metapopulation	The total population system that is composed of multiple local populations geographically separated but connected through dispersal.
Patch dynamics	A perspective that ecological systems are mosaics of patches, each exhibiting nonequilibrium dynamics and together determining the system-level behavior. Patches can be biotic or abiotic, ranging from a tree gap in a forest or a resource patch in a grassland to a whole ecosystem or a continent.
Pattern analysis	The procedures with which landscape pattern is quantified, primarily, using synoptic indices and spatial statistical methods.
Scale	The spatial or temporal dimension of a phenomenon. In landscape ecology, scale usually refers to grain and extent. Grain is the finest spatial or temporal unit in a data set, within which homogeneity is assumed, whereas extent is the total spatial area or temporal duration of a study. Grain and resolution are two related but distinct concepts. In general, fine-grained analyses require high-resolution data,

	but high-resolution data, after rescaling or aggregation, can also be used for coarse-grained analyses.
Scaling	The translation of information between or across spatial and temporal scales or organizational levels.
Spatial heterogeneity	The combination of discrete and continuous variations of one or more variables in a landscape, which can be characterized as patchiness, gradients, or a mixture of both. Spatial heterogeneity varies with scale in space and time.
Spatially explicit models	Models that explicitly take account of the locations of processes in a two- or three-dimensional space so that the spatial arrangement of landscape elements matters.

Definition of the Subject

Landscapes are spatially heterogeneous areas characterized by a mosaic of patches that differ in size, shape, contents, and history. When spatial heterogeneity is considered, the explicit treatment of scale becomes necessary and hierarchies emerge. Landscape ecology is the science of studying and improving the relationship between spatial pattern and ecological processes on a multitude of scales and organizational levels. In a broad sense, landscape ecology represents both a field of study and a scientific paradigm. As a highly interdisciplinary and transdisciplinary enterprise, landscape ecology integrates biophysical and analytical approaches with humanistic and holistic perspectives across natural and social sciences. Landscape ecology was initially developed in Europe. With theoretical developments in spatial ecology and technological advances in remote sensing and geospatial information processing, landscape ecology became an internationally recognized field of study in the 1980s. The most salient characteristics of landscape ecology are its emphasis on the pattern-process relationship and its focus on broad-scale ecological and environmental issues. Key research topics in landscape ecology include ecological flows in landscape mosaics (e.g., movements of water, nutrients, plant propagules, animals, and other materials), land use and land cover change, scaling, understanding the relationship between landscape pattern metrics and ecological processes, and landscape conservation and sustainability.

Introduction

Landscape ecology is an interdisciplinary field that aims to understand and improve the relationship between spatial pattern and ecological processes on a range of scales [1]. Although the term appeared in the 1930s, landscape ecology was not a recognized scientific field of global scope until the 1980s when remote sensing data and computers became widely accessible to ecologists and geographers. The 1980s was also a time period when ecological ideas of spatial heterogeneity and nonequilibrium dynamics flourished, and when landscape ecology took roots in North America. Today, landscape ecology is a well-established field of study, with active participation of ecological, geographical, and social scientists from around the world.

Landscape ecology has been dominated by two schools of thought: the European perspective and the North American perspective. At the risk of oversimplification, the European landscape ecology perspective may be considered as being characterized by a more holistic, humanistic, and society-centered view, with a focus on user-inspired and solution-driven research. The North American landscape ecology perspective, on the other hand, has been dominated by a more analytical and biological ecology–centered view, with a focus on basic science-oriented and question-driven studies. Cautions must be exercised, however, to avoid overinterpretation of such dichotomous characterization [2]. The two perspectives are neither inclusive nor exclusive; they are not contradictory but complementary. There are, and should be, other approaches to landscape ecology. For example, one may argue for an Australian landscape ecology perspective that focuses on pragmatic and functional approaches, typically, tied with land management, restoration, and conservation issues (e.g., [3]).

Landscape ecology is now a well-established interdisciplinary field of study, which is evidenced by several characteristics. These include an evolving but identifiable system of concepts, theories, principles, methods, and applications, a hierarchy of professional organizations from the international association to local chapters, a flagship journal, Landscape Ecology (http://www.springerlink.com/content/0921-2973), the adoption in educational and training programs by major universities and research institutes around the world, and an increasing number of publications in main-stream scientific journals which indicate its recognized status as well as its expanding impacts on related disciplines.

In this entry, I focus on the key concepts, research topics, and quantitative methods in landscape ecology. A number of textbooks on landscape ecology are available where more details on the contents covered here can be found [4–8].

What is Landscape Ecology?

Diverse Concepts of Landscape

The term, “landscape,” is a key concept in a number of fields, from social to geographical and ecological sciences. With the rise of landscape ecology in the past several decades, the concept of landscape has achieved a prominent status in the interdisciplinary literature. However, because of the plurality of its origins and interpretations, landscape has acquired various connotations. For example, the same word may refer to a natural landscape, a cultural landscape, a political landscape, an economic landscape, a mental landscape, an adaptive landscape, a landscape view, landscaping, or landscape painting [9, 10].

Even within the field of landscape ecology, the word, "landscape," has different meanings, and the differences usually hinge on the spatial scale and the contents of a landscape. For example, landscape has been defined as a kilometers-wide geographic area [11, 12], which corresponds to the "human-scale" landscape. This is the scale at which the field of landscape ecology was originally developed in Europe, and at which most landscape studies have been conducted around the world ever since. The human-scale landscape, in general, seems to coincide well with geographic units such as watersheds and urban regions [4], as well as spatial domains of human perception [13]. Thus, it resonates with the public, the decision makers, and researchers who are conscious about the environmental setting in which they live, work, and engage in recreation.

Many other landscape ecologists, however, have treated landscape as a multi-scale or hierarchical concept, meaning that a landscape is a spatially heterogeneous area that may be of various sizes, depending on the subject of study and the research questions at hand [6, 14, 15]. In this case, landscape is an "ecological criterion" [14], and its essence does not lie in its absolute scale but in its internal heterogeneity. Different plant and animal species perceive, experience, and respond to spatial heterogeneity at different scales, and patterns and processes in landscapes tend to have different characteristic scales [16]. Thus, a hierarchical concept of landscape, of course also encompassing the human-scale, is both sensible and necessary. Apparently, one does not need to consider a landscape of tens of square kilometers to study how grassland vegetation pattern affects the movement of beetles [17] or is affected by gophers [18].

The elements that constitute a landscape vary greatly in landscape ecological research. For simplicity, the components of a landscape may be classified as tangible versus intangible and biophysical versus cultural. This is not intended to represent a dichotomous view, but rather a continuum within which a variety of components coexist. Tress and Tress [10] proposed a "transdisciplinary landscape concept" that encompasses five dimensions: (1) landscape as a spatial entity, (2) landscape as a mental entity, (3) landscape as a temporal dimension, (4) landscape as a nexus of nature and culture, and (5) landscape as a complex system. Landscape ecological studies often have focused on some but not all of these dimensions. The concept of landscape provides a meeting ground for a number of disciplines, including archaeology, ecology, geography, geology, history, landscape architecture, and regional economics. To achieve its interdisciplinary and transdisciplinary goals, landscape ecology needs to appreciate and integrate the multifaceted perspectives on the culture-nature/people-place relationships that are offered by these diverse disciplines.

Evolving Concepts of Landscape Ecology

The definitions of landscape ecology are also diverse, although they are not quite as numerous as those of landscape (Table 11.1). Images can be powerfully inspiring, and this is especially true to someone who has a special interest in landscape patterns. Partly inspired by the conspicuous spatial patterns revealed in aerial photographs, the German geographer and botanist Carl Troll [19] coined the term "landscape ecology" and defined it later as "the study of the main complex causal relationships between the life communities and their environment in a given section of a landscape" [20, 21]. Carl Troll's training and research in multiple disciplines endowed him with the abilities to synthesize across, and innovate at the interface between, different fields. He was trained as a botanist; did his doctoral dissertation in plant physiology; and then spent decades working on the climatic, geologic, geographical, and ecological aspects of various landscapes in Europe, South America, and Africa. It is not difficult to understand why Troll could simultaneously appreciate the then-new idea of "ecosystem" put forward by Arthur Tansley [28], as well as the great potential for geospatial analysis presented by aerophotography. As a result of his attempt to integrate the "vertical" ecological approach with the "horizontal" geographical approach, a new field of study was born.

In the past several decades, landscape ecology has acquired a number of definitions which all are, in some way, related to Carl Troll's original definition. For example, Zonneveld [22] defined landscape ecology as "an aspect of geographical study which considers the landscape as a holistic entity, made up of different elements, all influencing each other." He advocated that the landscape should be studied as the "total character of a region," not "in terms of the separate aspects of its component elements" [22, 29]. This holistic landscape perspective continues and culminates in the work by Naveh [30], who described landscape ecology as the study of "the total spatial and functional entity of natural and cultural living space."

Some key ideas of contemporary landscape ecology, such as patch dynamics [31–33] and the patch-corridor-matrix model [11, 12], began to emerge in North America in the late 1970s, apparently with little connection to the European root. The early ideas of landscape ecology in North America were inspired by the theory of island biogeography [34], with an explicit focus on spatial heterogeneity. The first major communication between North American and European landscape ecologists occurred in 1981 when five American ecologists attended the first International Congress on Landscape Ecology in the Netherlands. Two years later, 25 ecologists (23 Americans, 1 Canadian, and 1 French) gathered at Allerton Park, Illinois of USA, to discuss the nature and future directions of landscape ecology. The report of this historic work, published in the following year [24], became an important guide to the incipient landscape ecologists in North America [35].

Why was such discussion necessary after landscape ecological research had been practiced for more than 40 years in Europe? The answer seems clear from Forman [36]: "What theory explains the spatial heterogeneity of energy, nutrients, water, plants, and animals at the level of a landscape, the setting in which we live?

Table 11.1 A list of definitions of landscape ecology

Definition	Source
The German geographer Carl Troll coined the term "landscape ecology" in 1939, and defined it in 1968 as "the study of the main complex causal relationships between the life communities and their environment in a given section of a landscape. These relationships are expressed regionally in a definite distribution pattern (landscape mosaic, landscape pattern) and in a natural regionalization at various orders of magnitude" (Troll 1968; cited in Troll 1971)	• Troll [19] • Troll [20] • Troll [21]
"Landscape ecology is an aspect of geographical study which considers the landscape as a holistic entity, made up of different elements, all influencing each other. This means that land is studied as the 'total character of a region', and not in terms of the separate aspects of its component elements" (Zonneveld 1972)	• Zonneveld [22]
"Landscape ecology is a young branch of modern ecology that deals with the interrelationship between man and his open and built-up landscapes" based on general systems theory, biocybernetics, and ecosystemology (Naveh and Liberman 1984). "Landscapes can be recognized as tangible and heterogeneous but closely interwoven natural and cultural entities of our total living space," and landscape ecology is "a holistic and transdisciplinary science of landscape study, appraisal, history, planning and management, conservation, and restoration" (Naveh and Liberman 1994)	• Naveh and Lieberman [5] • Naveh and Lieberman [23]
"A landscape is a kilometers-wide area where a cluster of interacting stands or ecosystems is repeated in similar form; landscape ecology, thus, studies the structure, function and development of landscapes" (Forman 1981). Landscape structure refers to "the spatial relationships among the distinctive ecosystems;" landscape function refers to "the flows of energy, materials, and species among the component ecosystems;" and landscape change refers to "the alteration in the structure and function of the ecological mosaic over time" (Forman and Godron 1986).	• Forman [11] • Forman [12]
"Landscape ecology focuses explicitly upon spatial pattern. Specifically, landscape ecology considers the development and dynamics of spatial heterogeneity, spatial and temporal interactions and exchanges across heterogeneous landscapes, influences of spatial heterogeneity on biotic and abiotic processes, and management of spatial heterogeneity" (Risser et al. 1984). "Landscape ecology is not a distinct discipline or simply a branch of ecology, but rather is the synthetic intersection of many related disciplines that focus on the spatial-temporal pattern of the landscape" (Risser et al. 1984).	• Risser et al. [24]
"Landscape ecology emphasizes broad spatial scales and the ecological effects of the spatial patterning of ecosystems" (Turner 1989).	• Turner [25]
"Landscape ecology is the study of the reciprocal effects of the spatial pattern on ecological processes," and "concerns spatial dynamics (including fluxes of organisms, materials, and energy) and the ways in which fluxes are controlled within heterogeneous matrices" (Pickett and Cadenasso 1995).	• Pickett and Cadenasso [14]
"Landscape ecology investigates landscape structure and ecological function at a scale that encompasses the ordinary elements of human landscape experience: yards, forests, fields, streams, and streets" (Nassauer 1997).	• Nassauer [26]

(continued)

Table 11.1 (continued)

Definition	Source
Landscape ecology is "ecology that is spatially explicit or locational; it is the study of the structure and dynamics of spatial mosaics and their ecological causes and consequences" and "may apply to any level of an organizational hierarchy, or at any of a great many scales of resolution" (Wiens 1999).	• Wiens [27]
"Landscape ecology emphasizes the interaction between spatial pattern and ecological process, that is, the causes and consequences of spatial heterogeneity across a range of scales" (Turner et al. 2001). "Two important aspects of landscape ecology . . . distinguish it from other subdisciplines within ecology": "First, landscape ecology explicitly addresses the importance of spatial configuration for ecological processes" and "second, landscape ecology often focuses on spatial extents that are much larger than those traditionally studied in ecology, often, the landscape as seen by a human observer" (Turner et al. 2001).	• Turner [6]
"Landscape ecology is the science and art of studying and influencing the relationship between spatial pattern and ecological processes across hierarchical levels of biological organization and different scales in space and time."	• Wu and Hobbs [1]

Alas, none." To develop such a landscape theory, broader scales that encompass multiple ecosystems need to be considered, and horizontal interactions have to be a focus of study. Thus, Forman and Godron [11, 12] defined landscape ecology as the study of the structure (spatial relationships among the distinctive landscape elements), function (flows of energy, materials, and species among landscape elements), and dynamics (temporal change in landscape structure and function) of landscapes. The main theme of landscape ecology in North America, with an unmistakable focus on spatial heterogeneity, was set in Risser et al. [24]:

> Landscape ecology focuses explicitly upon spatial pattern. Specifically, landscape ecology considers the development and dynamics of spatial heterogeneity, spatial and temporal interactions and exchanges across heterogeneous landscapes, influences of spatial heterogeneity on biotic and abiotic processes, and management of spatial heterogeneity.

Is landscape ecology a subdiscipline of ecology? The term itself apparently suggests that it is. Many ecologists do consider landscape ecology as a branch of ecology (e.g., [6]), and most ecology programs of major research universities worldwide now offer courses in landscape ecology. On the other hand, Zonneveld [22] indicated that landscape ecology is not part of biological sciences, but a branch of geography. Risser et al. [24] contemplated three ways in which landscape ecology may be viewed: as an intersection of many disciplines, as a separate discipline, or as a branch of ecology. They concluded that only the first option was "intellectually and practically the most persuasive." They further pointed out that "viewing landscape ecology as an interdisciplinary field of research avoids the issue of which discipline 'owns' landscape ecology" (a problem that may have hindered the healthy development of some interdisciplinary fields, such as human ecology, for which geography, sociology, and anthropology all have claimed ownership). The Allerton workshop report clearly recognized the importance of the multidimensionality of landscapes and the interdisciplinarity of landscape ecology:

> A major forcing function of landscapes is the activity of mankind, especially associated cultural, economic, and political phenomena. . . . Landscape ecology is not a distinct discipline or simply a branch of ecology, but rather is the synthetic intersection of many related disciplines that focus on the spatial-temporal pattern of the landscape" [24].

Today, a general consensus seems to have emerged that landscape ecology is not simply an academic discipline, but rather a highly interdisciplinary field of study [2, 37]. Landscape ecology is an interdisciplinary and transdisciplinary science that focuses on the relationship between spatial pattern and ecological processes across scales. The goal of landscape ecology is not only to understand this relationship but also to influence it so as to help achieve landscape sustainability [38–40]. As such, a pluralistic and hierarchical framework has been proposed to facilitate synergistic interactions between biophysical/pattern-process and holistic/humanistic perspectives (Fig. 11.1) [37, 38]. "Hierarchical" here refers to the varying degrees of interdisciplinary, the hierarchy of organizational levels, and the multiplicity of spatiotemporal scales of landscape ecological studies. "Pluralistic" indicates the necessity and importance of recognizing and valuing the different perspectives and methods in landscape ecology due to its diverse origins and goals.

Key Research Topics Landscape Ecology

Based on the suggestions by a group of leading landscape ecologists (Table 11.2), Wu and Hobbs [2] identified six key issues that characterize landscape ecology: (1) interdisciplinarity or transdisciplinarity, (2) integration between basic research and applications, (3) Conceptual and theoretical development, (4) education and training, (5) international scholarly communication and collaborations, and (6) outreach and communication with the public and decision makers. Wu and Hobbs [2] also summarized ten key research topics and priorities as follows:

1. Ecological flows in landscape mosaics: A primary goal of landscape ecology is to understand the reciprocal relationship between spatial pattern and ecological processes [14]. Understanding the mechanisms of flows of organisms, energy, material, and information in landscape mosaics is central to landscape ecology. In particular, the study of the effects of spatial pattern on population and ecosystem processes has made much progress in the past several decades. There is a need to integrate socioeconomic theory of landscape change into metapopulation models to make them more relevant to the issues of biodiversity conservation and landscape sustainability. The spread of invading species has become an increasingly important ecological and economic problem which deserves more research efforts.

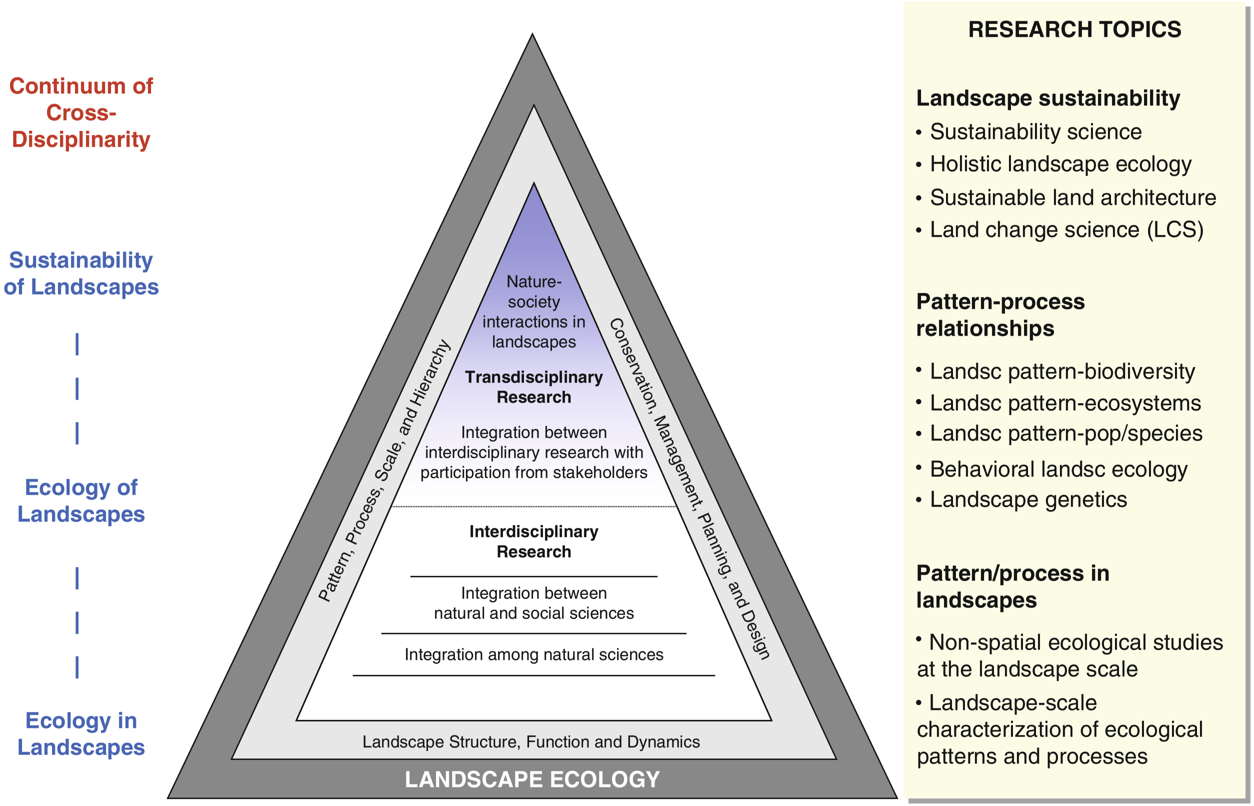


Fig. 1.1 A schematic representation of a hierarchical and pluralistic view of landscape ecology (Modified from [38])

Table 11.2 A list of major research topics in landscape ecology suggested by a group of leading landscape ecologists from around the world at the 16th Annual Symposium of the US Regional Association of the International Association for Landscape Ecology, held at Arizona State University, Tempe, in April 2001 [2]

Development of theory and principles	• Landscape mosaics and ecological flows • Land transformations • Landscape sustainability • Landscape complexity
Landscape metrics	• Norms or standards for metric selection, change detection, etc. • Integration of metrics with holistic landscape properties • Relating metrics to ecological processes • Sensitivity to scale change
Ecological flows in landscape mosaics	• Exchanges of organisms, material, energy, and information across the landscape • Effects of connectivity, edges, and boundaries • Spread of invading species • Spatial heterogeneity and ecosystem processes • Disturbances and patch dynamics
Optimization of landscape pattern	• Optimization of land use pattern • Optimal management • Optimal design and planning • New methods spatial optimization
Metapopulation theory	• Integration of the view of landscape mosaics • Incorporation of socioeconomic factors and management decisions
Scaling	• Extrapolating information across heterogeneous landscapes • Development of scaling theory and methods • Derivation of empirical scaling relations for landscape pattern and processes
Complexity and nonlinear dynamics of landscapes	• Landscapes as spatially extended complex systems • Landscapes as complex adaptive systems • Thresholds, criticality, and phase transitions • Self-organization in landscape structure and dynamics
Land use and land cover change	• Biophysical and socioeconomic drivers and mechanisms • Ecological consequences and feedbacks • Long-term landscape changes driven by economies and climate changes
Spatial heterogeneity in aquatic systems	• The relationship between spatial pattern and ecological processes in lakes, rivers, and oceans • Terrestrial and aquatic comparisons
Landscape-scale experiments	• Experimental landscape systems • Field manipulative studies • Scale effects in experimental studies
New methodological developments	• Integration among observation, experimentation, and modeling • New statistical and modeling methods for spatially explicit studies • Interdisciplinary and transdisciplinary approaches
Data collection and accuracy assessment	• Multiple-scale landscape data • More emphasis on collecting data on organisms and processes • Data quality control • Metadata and accuracy assessment

(continued)

Table 11.2 (continued)

Fast changing and chaotic landscapes	• Rapidly urbanizing landscapes • War zones • Other highly dynamic landscapes
Landscape sustainability	• Developing operational definitions and measures that integrate ecological, social, cultural, economic, and aesthetic components • Practical strategies for creating and maintaining landscape sustainability
Human activities in landscapes	• The role of humans in shaping landscape pattern and processes • Effects of socioeconomic and cultural processes on landscape structure and functioning
Holistic landscape ecology	• Landscape ecology as an anticipative and prescriptive environmental science • Development of holistic and systems approaches

2. Causes, processes, and consequences of land use and land cover change: Land use and land cover change is arguably the most important driver for changes in the structure and function of landscapes. Land use and land cover change is driven primarily by socioeconomic forces, and is one of the most important and challenging research areas in landscape ecology. Numerous studies have been carried out to investigate the effects of land use and land cover change on biodiversity and ecological flows in human-dominated landscapes. More research efforts are needed to incorporate the insights of economic geography which studies how economic activity is distributed in space and resource economics which determines how land will be used [41]. Long-term landscape changes induced by economic activities and climate change, as well as "land use legacies" (i.e., the types, extents, and durations of persistent effects of prior land use on ecological patterns and processes) need to be emphasized in future research.
3. Nonlinear dynamics and landscape complexity: Landscapes are spatially extended complex systems which exhibit emergent properties, phase transitions, and threshold behavior. To understand the complexity of landscapes, concepts and methods from the science of complexity and nonlinear dynamics should be helpful. For example, self-organization, percolation theory, complex adaptive systems (CAS), fractal geometry, cellular automata, and genetic algorithms have been used in the study of spatiotemporal dynamics of landscapes (e.g., [42–46]). However, the theoretical potential and practical implications of these concepts and methods are yet to be fully explored.
4. Scaling: Scaling refers to the translation of information from one scale to another across space, time, or organizational levels. Spatial scaling, in particular, is essential in both the theory and practice of landscape ecology because spatial heterogeneity does not make any sense without the consideration of scale [47]. While scale effects are widely recognized in landscape ecology, scaling-up or scaling-down across heterogeneous landscapes remains a grand challenge in landscape ecology and beyond [48]. General rules and pragmatic

methods for scaling landscape patterns and processes need to be developed and tested.

5. Methodological advances: Landscape variables are often spatially autocorrelated and spatially dependent, which poses serious challenges for using traditional statistical methods based on the assumption of independence of observations. The spatial autocorrelation and dependence that traditional statistical methods try to get rid of are usually what landscape analyses intend to get at. Thus, spatial statistical methods that directly deal with spatial autocorrelation and dependence have increasingly been used in landscape ecology. Also, most landscape ecological problems need to be studied over large and multiple scales in a spatially explicit manner. This need poses problems such as the lack of replicability or "pseudoreplication" [49]. To get to the processes and mechanisms of landscape phenomena, landscape ecology has developed a suite of spatially explicit modeling approaches [50, 51]. In both spatial analysis and modeling of landscapes, remote sensing and GIS (geographic information systems) have become indispensable.
6. Relating spatial pattern measures to ecological processes: To understand the relationship between pattern and process, quantifying landscape pattern is necessary. Indeed, landscape pattern analysis has been a major part of landscape ecological research for the last few decades. A number of landscape metrics (Table 11.3) and spatial statistical methods have been developed and applied for describing and comparing the spatial patterns of landscapes, monitoring and predicting changes in landscape patterns, and relating spatial pattern to ecological processes at a particular scale or across a range of scales [47, 53, 54]. Nevertheless, a sound ecological understanding of these spatial analysis methods is yet to be fully developed [55].
7. Integrating humans and their activities into landscape ecology: Socioeconomic processes are the primary drivers for land use and land cover change which in turn determines the structure, function, and dynamics of most landscapes. Social and economic processes have increasingly been integrated into landscape ecological studies. The need for incorporating humans, including their perceptions, value systems, cultural traditions, and socioeconomic activities, into landscape ecology has made it a highly interdisciplinary and transdisciplinary enterprise [38, 56]. That said, effectively integrating human-related processes into ecology may remain one of the ultimate challenges for landscape ecologists in years to come.
8. Optimization of landscape pattern: If spatial pattern significantly influences ecological processes in the landscape, then there must be certain patterns that are better than others in terms of promoting ecosystem functioning and services. This is a question of landscape pattern optimization (e.g., optimization of land use pattern, optimal landscape management, optimal landscape design, and planning). For example, can landscape patterns be optimized in terms of both the composition and configuration of patches and matrix characteristics to maximize biodiversity and ecosystem services? Are there optimal ways of "spatially meshing nature and culture" to promote landscape

Table 11.3 Some commonly used landscape metrics [4, 52]

Landscape metric	Abbreviation	Description
Patch shape index	PSI	A measure for the complexity of the shape of a given patch: $PSI = \frac{P}{2\sqrt{A\pi}}$ where P is the perimeter of a patch, and A is the area of the patch. $PSI = 1$ for circles; $PSI = 1.1283$ for squares; and $PSI = 1.1968$ for a rectangle (2 L by L). $1/D$ is called “compactness” (see Forman [4]).
Perimeter/area ratio	PAR	A measure of the complexity of the shape of a patch: $PAR = P/A$ where P is the perimeter of a patch, and A is the area of the patch.
Number of patches	NP	The total number of patches in the landscape.
Patch density	PD	The number of patches per square kilometer (i.e., 100 ha).
Total edge	TE	The sum of the lengths of all edge segments (unit: meter).
Edge density	ED	The total length of all edge segments per hectare for the class or landscape of consideration (unit: m/ha).
Patch richness	PR	The number of different patch types in the landscape.
Patch richness density	PRD	The number of patch types per square kilometer (or 100 ha).
Shannon’s diversity index	SHDI	A measure of patch diversity in a landscape that is determined by both the number of different patch types and the proportional distribution of area among patch types: $H = -\sum_{i=1}^{m} p_i \ln(p_i)$ where m is the total number of patch types and p_i is the proportion of the landscape area occupied by patch type i.

Dominance index	D	A measure of the degree of dominance by one or a few patch types in a landscape: $D = H_{\max} + \sum_{i=1}^{m} p_i \ln p_i$ where $H_{\max}$ is the maximum diversity when all patch types are present in equal proportions, m is the total number of patch types, and p_i is the proportion of the landscape area occupied by patch type i. Small values of D tend to indicate landscapes with numerous land use types of similar proportions.
Largest patch index	LPI	The ratio of the area of the largest patch to the total area of the landscape (unit: percentage).
Mean patch size	MPS	The average area of all patches in the landscape (unit: ha).
Patch size standard deviation	PSSD	The standard deviation of patch size in the entire landscape (unit: ha).
Patch size coefficient of variation	PSCV	The standard deviation of patch size divided by mean patch size for the entire landscape (unit: percentage).
Landscape shape index	LSI	A modified perimeter/area ratio of the form: $LSI = \frac{0.25E}{\sqrt{A}}$ where E is the total length of patch edges and A is the total area of the landscape (unitless).
Mean patch shape index	MSI	A patch-level shape index averaged over all patches in the landscape: $MSI = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\frac{0.25P_{ij}}{\sqrt{a_{ij}}}\right]}{N}$ where P_{ij} and a_{ij} are the perimeter and area of patch ij, respectively, and N is the total number of patches in the landscape (unitless).
Area-weighted mean patch shape index	AWMSI	Mean patch shape index weighted by relative patch size: $AWMSI = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left(\frac{0.25P_{ij}}{\sqrt{a_{ij}}}\right)\left(\frac{a_{ij}}{A}\right)\right]$ where P_{ij} and a_{ij} are the perimeter and area of patch ij, respectively, A is the total area of the landscape, m is the number of patch types, and n is the total number of patches of type i (unitless).

(continued)

Table 11.3 (continued)

Landscape metric	Abbreviation	Description
Double-log fractal dimension	DLFD	The fractal dimension for the entire landscape which is equal to 2 divided by the slope of the regression line between the logarithm of patch area and the logarithm of patch perimeter: $DLFD = \frac{2}{\left[\frac{\left[N\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\ln\left(P_{ij}\right)\ln\left(a_{ij}\right)\right)\right]-\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln\left(a_{ij}\right)\right)\right]}{\left(N\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\ln\left({P_{ij}}^{2}\right)\right)\right)-\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln\left(P_{ij}\right)\right)^{2}}\right]}$ where P_{ij} and a_{ij} are the perimeter and area of patch ij, respectively, m is the number of patch types, n is the total number of patches of type i, and N is the total number of patches in the landscape (unitless).
Mean patch fractal dimension	MPFD	The average fractal dimension of individual patches in the landscape, which is the summation of fractal dimension for all patches divided by the total number of patches in the landscape: $FD = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{2\ln\left(0.25P_{ij}\right)}{\ln\left(a_{ij}\right)}\right)}{N}$ where P_{ij} and a_{ij} are the perimeter and area of patch ij, respectively, m is the number of patch types, n is the total number of patches of type i, and N is the total number of patches in the landscape (unitless).

Area-weighted mean patch fractal dimension	AWMFD	The patch fractal dimension weighted by relative patch area: $AWMPED = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{2\ln\left(0.25P_{ij}\right)}{\ln\left(a_{ij}\right)}\left(\frac{a_{ij}}{A}\right)\right)$ where P_{ij} and a_{ij} are the perimeter and area of patch ij, respectively, m is the number of patch types, n is the total number of patches of type i, and A is the total area of the landscape (unitless).
Contagion	CONT	An information theory–based index that measures the extent to which patches are spatially aggregated in the landscape [57]: $CONT = \left\lfloor 1 + \sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij}\ln\left(p_{ij}\right)/2\ln(m)\right\rfloor (100)$ where p_{ij} is the probability that two randomly chosen adjacent pixels belong to patch type i and j, m is the total number of patch types in the landscape (unitless).

sustainability? These are some of the challenging questions that landscape ecologists ought to address now and in the future. Spatial optimization of landscape pattern for environmental purposes presents exciting research opportunities and requires interdisciplinary approaches.

9. Landscape conservation and sustainability: Biodiversity, ecosystem functions, and human activities, all take place in landscapes. Landscape fragmentation profoundly alters ecological and socioeconomic processes. Thus, the importance of applying landscape ecological principles in biodiversity conservation and sustainable development has been increasingly recognized. However, specific landscape ecological guidelines for biodiversity conservation are needed, and a comprehensive and operational definition of landscape sustainability is yet to be developed.
10. Data acquisition and accuracy assessment: Landscape ecological studies use large-scale and multi-scale data. A suite of advanced technologies are readily available, including various remote sensing techniques, GIS, GPS (global positioning systems), and spatial analysis and modeling approaches. However, ecological understanding of species and ecosystems is essential in landscape ecology, and this requires the collection of basic biological data of landscapes. Also, to ensure the quality of landscape data, error analysis, uncertainty analysis, and accuracy assessment have become a key issue in landscape ecological research.

Future Directions

Landscape ecology is a highly interdisciplinary field of study which is characterized, most conspicuously, by its spatial explicitness in dealing with ecological problems in theory and practice. Emphasis on spatial heterogeneity begs questions of the pattern-process relationships and scale. Studying spatial pattern without relating it to ecological processes is superficial, and investigating ecological processes without consideration of spatial pattern is incomplete. From this perspective, landscape ecology is a science of heterogeneity and scale, providing a new scientific paradigm for ecology and other related fields.

On the other hand, with increasing human dominance in the biosphere, emphasis on broad spatial scales makes it inevitable to deal with humans and their activities. As a consequence, humanistic and holistic perspectives have been and will continue to be central in landscape ecological research. Thus, landscape ecology has become increasingly relevant to sustainability research and practice [38, 56]. First, landscape ecology provides a hierarchical and integrative ecological basis for dealing with issues of biodiversity and ecosystem functioning from fine to broad scales. Second, landscape ecology has already developed a number of holistic and humanistic approaches to studying nature–society interactions. Third, landscape ecology offers theory and methods for studying the effects of spatial configuration of biophysical and socioeconomic component on the sustainability of a place. Fourth, landscape ecology has developed a suite of pattern metrics and indicators which can be used for quantifying sustainability in a geospatially explicit manner. Finally, landscape ecology provides both theoretical and methodological tools for dealing with scaling and uncertainty issues that are fundamental to most nature–society interactions.

To move forward, future landscape ecological studies need to further address the key research topics as discussed earlier in this entry. In addition, concerted efforts need to be made to focus on sustainability-related research questions. For example, what theories, principles, and methods of landscape ecology are pertinent to sustainability and how can they be operationalized? How does landscape pattern or spatial heterogeneity affect sustainability? How do ecological, economic, and social patterns and processes in landscapes change with scale and interact to influence sustainability? How is landscape sustainability measured and what roles can landscape metrics play in all this? How can landscape models to project sustainability trajectories in response to environmental, economic, social, and institutional changes be developed? And finally, how can landscape ecology help design sustainable landscapes?

Bibliography

Primary Literature

1. Wu J, Hobbs R (2007) Landscape ecology: the-state-of-the-science. In: Wu J, Hobbs R (eds) Key topics in landscape ecology. Cambridge University Press, Cambridge, pp 271–287
2. Wu J, Hobbs R (2002) Key issues and research priorities in landscape ecology: an idiosyncratic synthesis. Landscape Ecol 17:355–365
3. Ludwig J, Tongway D, Freudenberger D, Noble J, Hodgkinson K (1997) Landscape ecology, function and management: principles from Australia's rangelands. CSIRO, Collingwood
4. Forman RTT (1995) Land mosaics: the ecology of landscapes and regions. Cambridge University Press, Cambridge
5. Naveh Z, Lieberman AS (1984) Landscape ecology: theory and application. Springer, New York
6. Turner MG, Gardner RH, O'Neill RV (2001) Landscape ecology in theory and practice: pattern and process. Springer, New York
7. Burel F, Baudry J (2003) Landscape ecology: concepts, methods and applications. Science, Enfield
8. Farina A (1998) Principles and methods in landscape ecology. Chapman & Hall, London
9. Mitchell D (2000) Cultural geography: a critical introduction. Blackwell, Oxford
10. Tress B, Tress G (2001) Capitalising on multiplicity: a transdisciplinary systems approach to landscape research. Landscape Urban Plan 57:143–157
11. Forman RTT (1981) Interaction among landscape elements: a core of landscape ecology. In: Tjallingii SP, de Veer AA (eds) Perspectives in landscape ecology: contributions to research, planning and management of our environment. Pudoc, Wageningen, pp 35–48
12. Forman RTT, Godron M (1986) Landscape ecology. Wiley, New York
13. Gobster PH, Nassauer JI, Daniel TC, Fry G (2007) The shared landscape: what does aesthetics have to do with ecology? Landscape Ecol 22:959–972

14. Pickett STA, Cadenasso ML (1995) Landscape ecology: spatial heterogeneity in ecological systems. Science 269:331–334
15. Urban DL, O'Neill RV, Shugart HH (1987) Landscape ecology: a hierarchical perspective can help scientists understand spatial patterns. BioScience 37:119–127
16. Wu J, Loucks OL (1995) From balance-of-nature to hierarchical patch dynamics: a paradigm shift in ecology. Quart Rev Biol 70:439–466
17. Wiens JA, Milne BT (1989) Scaling of "landscape" in landscape ecology, or, landscape ecology from a beetle's perspective. Landscape Ecol 3:87–96
18. Wu J, Levin SA (1994) A spatial patch dynamic modeling approach to pattern and process in an annual grassland. Ecol Monogr 64(4):447–464
19. Troll C (1939) Luftbildplan and ökologische Bodenforschung. Zeitschrift der Gesellschaft für Erdkunde zu Berlin 7–8:241–298
20. Troll C (1968) Landschaftsökologie. In: Tuxen B (ed) Pflanzensoziologie und Landschaftsökologie, Berichte das 1963 internalen symposiums der internationalen vereinigung fur vegetationskunde. Junk, The Hague, pp 1–21
21. Troll C (1971) Landscape ecology (geoecology) and biogeocenology–a terminology study. Geoforum 8(71):43–46
22. Zonneveld IS (1972) Land evaluation and land(scape) science. International Institute for Aerial Survey and Earth Sciences, Enschede
23. Naveh Z, Lieberman AS (1994) Landscape ecology: theory and application, 2nd edn. Springer, New York
24. Risser PG, Karr JR, Forman RTT (1984) Landscape ecology: directions and approaches. Illinois Natural History Survey Special Publication 2, Champaign
25. Turner MG (1989) Landscape ecology: the effect of pattern on process. Ann Rev Ecol Syst 20:171–197
26. Nassauer JI (1997) Culture and landscape ecology: insights for action. In: Nassauer JI (ed) Placing nature. Island Press, Washington, DC, pp 1–11
27. Wiens JA (1999) Toward a unified landscape ecology. In: Wiens JA, Moss MR (eds) Issues in landscape ecology. International Association for Landscape Ecology, Snowmass Village, pp 148–151
28. Tansley AG (1935) The use and abuse of vegetational concepts and terms. Ecology 16:284–307
29. Zonneveld IS (1989) The land unit – a fundamental concept in landscape ecology, and its applications. Landscape Ecol 3:67–86
30. Naveh Z (1991) Some remarks on recent developments in landscape ecology as a transdisciplinary ecological and geographical science. Landscape Ecol 5:65–73
31. Burgess RL, Sharpe DM (eds) (1981) Forest Island dynamics in man-dominated landscapes. Springer, New York
32. Levin SA, Paine RT (1974) Disturbance, patch formation and community structure. Proc Nat Acad Sci USA 71:2744–2747
33. Pickett STA, Thompson JN (1978) Patch dynamics and the design of nature reserves. Biol Conser 13:27–37
34. MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton
35. Wiens JA (2008) Allerton Park 1983: the beginnings of a paradigm for landscape ecology? Landscape Ecol 23:125–128
36. Forman RTT (1983) An ecology of the landscape. BioScience 33:535
37. Wu J, Hobbs RJ (eds) (2007) Key topics in landscape ecology. Cambridge University Press, Cambridge
38. Wu J (2006) Landscape ecology, cross-disciplinarity, and sustainability science. Landscape Ecol 21:1–4
39. Wu J (2008) Making the case for landscape ecology: an effective approach to urban sustainability. Landscape J 27:41–50

40. Wu J (2010) Urban sustainability: an inevitable goal of landscape research. Landscape Ecol 25:1–4
41. O'Neill RV (1999) Theory in landscape ecology. In: Wiens JA, Moss MR (eds) Issues in landscape ecology. International Association for Landscape Ecology, Snowmass Village, pp 1–5
42. With KA, Pavuk DM, Worchuck JL, Oates RK, Fisher JL (2002) Threshold effects of landscape structure on biological control in agroecosystems. Ecol Appl 12:52–65
43. Milne BT (1998) Motivation and benefits of complex systems approaches in ecology. Ecosystems 1:449–456
44. Levin SA (1998) Ecosystems and the biosphere as complex adaptive systems. Ecosystems 1:431–436
45. Levin SA (2005) Self-organization and the emergence of complexity in ecological systems. BioScience 55:1075–1079
46. Gardner RH, Milne BT, Turner MG, O'Neill RV (1987) Neutral models for the analysis of broad-scale landscape pattern. Landscape Ecol 1:19–28
47. Wu J (2004) Effects of changing scale on landscape pattern analysis: scaling relations. Landscape Ecol 19:125–138
48. Wu J, Jones KB, Li H, Loucks OL (eds) (2006) Scaling and uncertainty analysis in ecology: methods and applications. Springer, Dordrecht
49. Hargrove WW, Pickering J (1992) Pseudoreplication: a sine qua non for regional ecology. Landscape Ecol 6:251–258
50. Turner MG, Gardner RH (1991) Quantitative methods in landscape ecology: the analysis and interpretation of landscape heterogeneity. Springer, New York
51. Wu J, Levin SA (1997) A patch-based spatial modeling approach: conceptual framework and simulation scheme. Ecol Model 101:325–346
52. McGarigal K, Marks BJ (1995) FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Genernal Technnical Report PNW-GTR-351. Pacific Northwest Research Station, USDA-Forest Service, Portland
53. Gustafson EJ (1998) Quantifying landscape spatial pattern: what is the state of the art? Ecosystems 1:143–156
54. Fortin M-J, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge
55. Li HB, Wu JG (2004) Use and misuse of landscape indices. Landscape Ecol 19:389–399
56. Naveh Z (2007) Landscape ecology and sustainability. Landscape Ecol 22:1437–1440
57. Li H, Reynolds JF (1993) A new contagion index to quantify spatial patterns of landscapes. Landscape Ecol 8(3):155–162

Books and Reviews

Barrett GW, Peles JD (eds) (1999) Landscape ecology of small mammals. Springer, New York

Bissonette JA (ed) (1997) Wildlife and landscape ecology: effects of pattern and scale. Springer, New York

Dramstad WE, Olson JD, Forman RTT (1996) Landscape ecology principles in landscape architecture and land-use planning. Harvard University Graduate School of Design/Island Press, Cambridge

Hansson L, Fahrig L, Merriam G (1995) Mosaic landscapes and ecological processes. Chapman & Hall, London

Ingegnoli V (2002) Landscape ecology: a widening foundation. Springer, Berlin

Nassauer JI (ed) (1997) Placing nature: culture and landscape ecology. Island Press, Washington, DC

Naveh Z, Lieberman AS (1994) Landscape ecology: theory and application. Springer, New York
Turner MG (2005) Landscape ecology: what is the state of the science? Ann Rev Ecol Systemat 36:319–344
Wiens J, Moss M (eds) (2005) Issues and perspectives in landscape ecology. Cambridge University Press, Cambridge
Wu J, Hobbs R (eds) (2007) Key topics in landscape ecology. Cambridge University Press, Cambridge
Wu J, Jones B, Li H, Loucks O (eds) (2006) Scaling and uncertainty analysis in ecology. Springer, Dordrecht

Chapter 12
Marine Biogeochemistry

Walker O. Smith Jr., Eileen E. Hofmann, and Anna Mosby

Glossary

Autotrophic	Organisms whose mode of nutrition is photosynthesis.
Biogeochemistry	The biological and chemical processes that transform and cycle elements over various time and space scales and that determine the composition of the environment.
Biological pump	The biological processes and transformations that move carbon from the surface to depth.
Cyanobacteria	Prokaryotic phytoplankton.
Diatom	Phytoplankton which are encased in frustule consisting of silica.
Euphotic zone	The surface layer of the ocean where most primary production occurs, generally considered to be the depth to which 1% of surface radiation penetrates.
Heterotrophic	Organisms who require reduced organic carbon as an energy and carbon source.
Nutrient	Element that is required for biological activity and growth.
Oxidation	Chemical reaction in which reactant loses electrons; half-reaction paired with reduction.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

W.O. Smith Jr. (✉) • A. Mosby
Department of Biological Sciences, Virginia Institute of Marine Science,
College of William & Mary, Gloucester Pt, VA 23062, USA
e-mail: wos@vims.edu; afmosby@gmail.com

E.E. Hofmann
Center for Coastal Physical Oceanography, Old Dominion University, Norfolk,
VA 23508, USA
e-mail: hofmann@ccpo.odu.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_12,

Photosynthesis	The process by which radiant energy from the sun is transformed into chemical energy that can later be used to reduce carbon dioxide to organic sugars, which in turn are coupled to biochemical pathways to produce all compounds necessary for cell growth.
Phytoplankton	Microscopic, often unicellular, floating autotrophs that live in the ocean's surface layer and form the base of nearly all marine food webs.
Reduction	Chemical reaction in which reactant gains electrons; half-reaction paired with oxidation.

Definition of the Subject and Its Importance

The biogeochemistry of the world oceans has been studied for many decades, and major advances in understanding have been linked with development of new techniques and tools that allow the accurate representation of various organic and inorganic pools within the water. The classic study of Redfield [1] showed that some critical bioactive compounds (carbon, nitrogen, phosphorus, oxygen) occur in particular ratios to one another that are relatively invariant over space and time and provided a description of the relationship between the ratio of nitrogen to phosphorus (N:P) for inorganic and plankton pools. The processes that control these compounds were assessed, and it was concluded that phosphorus concentrations are largely controlled by terrestrial inputs, whereas nitrogen is under biological control.

Subsequent studies have provided more detailed investigations of the processes controlling these ratios. These studies benefited from the development and standardization of methods for accurately measuring dissolved organic carbon (DOC) and dissolved organic nitrogen (DON). The improved methodology, mostly developed during the 1980s, allowed the spatial (vertical and horizontal) and temporal changes of both DOC and DON to be quantitatively described.

Recognition of the importance of the flux of organic carbon to depth in mediating the marine response to increased atmospheric carbon dioxide concentrations stimulated development of technical approaches and instruments for assessing and quantifying the biological pump. This component of marine biogeochemical cycles is still a poorly constrained component of numerical models developed for simulation of ocean carbon cycling and climate, and technological approaches that result in better assessment of the flux of organic matter to depth continue to be developed and refined. Also, numerical models of biogeochemical processes are providing insights into critical processes and provide frameworks that allow measurements to be projected over larger space and timescales. Continued measurement and modeling of oceanic biogeochemical cycles is essential for understanding and projecting responses to natural and anthropogenic-induced climate change.

Introduction

The ocean is the dominant surface feature that has controlled much of the evolution, distribution, and success of life on earth, and the changes in ocean chemistry reflect the interaction with biota throughout geological time. The oceans were originally anoxic, but the evolution of organisms with oxygen-generating processes (photosynthesis) resulted in the conversion of the oceans to an oxygenated environment, which greatly altered the availability of some elements for those organisms. The cycling of elements within the earth's oceans and the complex relationships among the biological, chemical, and geological processes are the core of the study of marine biogeochemistry. Understanding these relationships is difficult and is further complicated by the space and time variability of the dominant processes that control the cycling of the different elements. Understanding the interactions and linkages among and between the cycles of biogeochemical elements is critically important for assessing and projecting the nature, degree, and direction of changes in ocean processes that may result from changes induced by natural and/or anthropogenic activities.

Elements in the ocean have characteristic vertical and horizontal distributions that result from the processes that regulate their long-term source/sink relationships. For example, oceanic carbon dioxide (CO_2) distributions are characterized by a horizontal concentration gradient that increases from the equator to the poles, which results from the greater dissolution of CO_2 in colder water. Carbon dioxide concentrations generally increase with depth due to remineralization in the deeper, older waters relative to its removal at the surface. Other elements may be controlled by different factors (e.g., sources from the sediments or hydrothermal vents; atmospheric sources) and have different vertical and horizontal patterns, but all interact to create the observed vertical distributions in the ocean. Understanding marine biogeochemistry requires knowledge not only of specific processes regulating a particular element, but also an understanding of the interdisciplinary aspects that control these cycles.

Nutrients are the biogeochemical elements that are required for biological activity. Some elements are greatly reduced in their concentrations by chemical or biological processes and can reach such low concentrations that they subsequently limit the growth of organisms in the sea. Such elements are thought of as limiting nutrients in the sense of the German agricultural chemist, Justus von Liebig, who suggested that the growth of plants is limited not by the total amount of resources, but by the resource in lowest abundance relative to the others.

Plant growth in the ocean is known to be limited by a small number of nutrients that include nitrogen, phosphorus, iron, silicic acid, and inorganic carbon. The cycling and processes that control the concentrations of these limiting nutrients are critical in the regulation of carbon cycling in the ocean, and hence their study forms the basis for most biogeochemical research.

The biogeochemical cycles described in subsequent sections use carbon as a "common denominator." Carbon is the basic component of organic matter, and with the advent of industrialization is being added to the atmosphere at an unprecedented, rapid rate, which is changing atmospheric temperatures and impacting the thermal equilibrium of the ocean. Also, carbon is absorbed from the atmosphere at the ocean surface where it reacts with ocean water to produce carbonic acid, thereby making ocean waters more acidic (reducing the pH), which has profound impacts on oceanic chemistry and biological activity. Thus, the production and oxidation of organic matter in the ocean has numerous critical interactions with all other elemental cycles, and is a major regulator of all marine biogeochemical cycles.

The Biogeochemical Cycle of Carbon

Carbon is the primary building block for all life because of its chemical ability to form a myriad of covalent bonds with itself and numerous other elements. As a result, the numerous complex organic compounds that form the basis of life systems are based on carbon. In the present-day ocean, synthesis of organic molecules (photosynthesis) is done largely by phytoplankton, which converts the radiant energy from the sun into chemical energy in the form of adenosine triphosphate (ATP). The ATP, along with reductant, is used to reduce CO_2 into simple sugars, which are in turn modified into all of the compounds required for cellular metabolism, growth, and division. Photosynthesis is dependent on energy from the sun, thereby confining this process to the euphotic zone, which is the part of the upper water column that receives at least 1% of the irradiance that reaches the sea surface. Phytoplankton require energy for the uptake and assimilation of nearly all elements. This dependence on light generally results in vertical distributions of nutrients that are characterized by reduced concentrations in the euphotic zone, where photosynthesis and growth are most active, and increased concentrations at depth, where photosynthesis and growth are reduced or absent (Fig. 12.1). This vertical profile is a typical of nutrient distributions throughout the oceans. The organic matter generated by photosynthesis and growth has roughly an inverse relationship to that of the inorganic building blocks (Fig. 12.1).

Redfield [1] suggested that organic matter (carbon, C) production in the sea occurs in relatively constant elemental ratios given by the relationship:

$$\begin{gathered}106CO_2 + 16H^+ + 16NO_3^- + H_3PO_4 + 122H_2O \\ \leftrightarrow (CH_2O)_{106}(NH_3)_{16}(H_3PO_4) + 138O_2\end{gathered}$$

This relationship describes the reaction of CO_2 with hydrogen (H), nitrate (NO_3^-), phosphate (H_3PO_4), and water (H_2O) within the photosynthetic process to produce

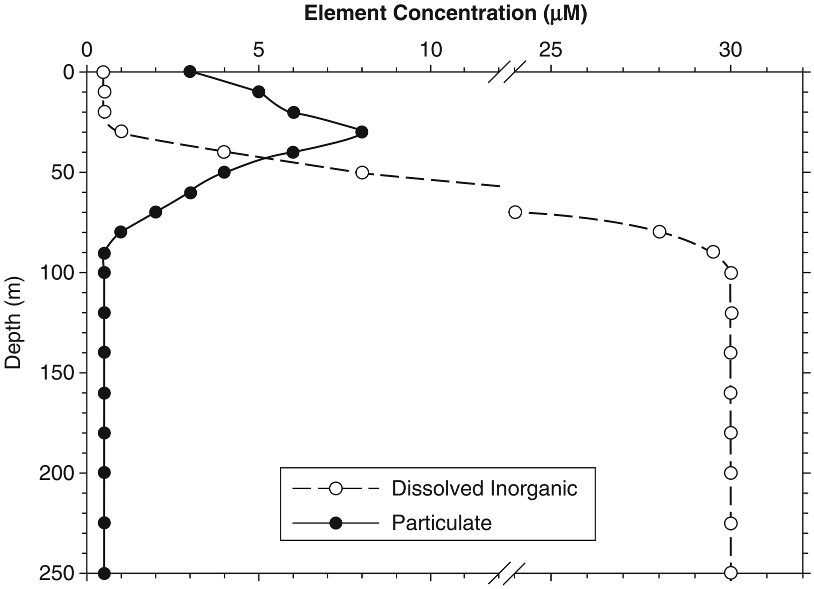


Fig. 12.1 Generalized *vertical* distributions of dissolved inorganic elements and particulate matter produced by phytoplankton photosynthesis. Particulate matter concentrations are less than those expected from the disappearance of inorganic elements because of removal by various processes to depth (see Fig. 12.2). Similarly, the particulate matter *vertical* distribution is less uniform because the time-scale of redistribution of particles is much faster than that of the inorganic elements

($\leftrightarrow$) organic carbon-nitrogen-phosphorus compounds $\left[(CH_2O)_{106}(NH_3)_{16}H_3PO_4\right]$ and gaseous oxygen (O_2). The numbers preceding the compounds indicate the amount of each. The relationship is reversible ($\leftrightarrow$) because metabolism (oxidation) of the organic matter produced by photosynthesis regenerates inorganic C, nitrogen (N), and phosphorus (P) in the same ratio and utilizes oxygen. The C:N:P ratio of 106:16:1 obtained from the above relationship is a basic paradigm of marine biogeochemistry. However, Redfield recognized that the C:N:P ratios vary within plankton types and with time, a fact that has been further established in more recent studies. The departure from the basic ratio provides insights in how marine ecosystems change and/or adapt to modified environmental or biological conditions.

All marine organisms contribute to the carbon cycle by moving carbon between organic and inorganic forms, but some marine organisms are able to use calcification to transform inorganic carbon, using bicarbonate and dissolved calcium from the water column to produce calcium carbonate ($CaCO_3$), which is then used to form a skeleton or protective shell [2]. The dissolution of calcium carbonate back into its original components is one of the primary means by which the particulate components reenter the water column, keeping the inorganic carbon cycle running. Although some of the calcium carbonate dissolved back into the water column comes from dead organisms, a large portion is contributed by phytoplankton from coccolithophorids, the genus coccolithophorid, which produce and shed calcium carbonate shells, making them a major contributor to the inorganic carbon cycle [2]. The calcium carbonate not immediately dissolved back into the water column is removed by sinking, with coccolithophorids comprising a major component of the carbon found in marine sediments. A by-product of calcification is CO_2, which either remains in the water column or reenters the biological pump through photosynthesis [2].

The importance of iron (Fe) and silicic acid ($Si(OH)_4$) in regulating carbon production in oceanic systems has also been established. Iron is required by all living organisms for a variety of metabolic processes, and silicon (Si) is needed by an important phytoplankton functional group, the diatoms, which are characterized by a hard silica shell. Diatoms remove silicic acid in approximately a 1:1 ratio to N, and the P:Fe ratio is approximately 1,000:1. Both ratios show considerable plasticity and their uptake ratios are related to other environmental variables as well [3, 4].

The organic material produced in the upper water column via photosynthesis is used by heterotrophic organisms (e.g., bacteria, zooplankton) and transformed by their metabolism and growth processes. The unassimilated ingestion of these organisms (fecal pellet production) sinks and is oxidized below the euphotic zone by a host of heterotrophic organisms (from bacteria to ciliates to scavenging, mobile animals), thereby converting the organic matter to CO_2. Also, particle aggregates formed from phytoplankton cells, detritus, and dead organisms sink from the euphotic zone and are oxidized. The unidirectional movement of large particles to depth and their remineralization defines the biological pump (Fig. 12.2), which also contributes to the generation of "nutrient-like" profiles in the ocean. The processes that contribute to the fluxes within the biological pump are critical to understanding the marine carbon cycle.

Atmospheric fluxes of CO_2 into and out of the ocean vary spatially. In general, equatorial waters tend to be large sources of CO_2 (net fluxes are from the ocean to the atmosphere). The equatorial Pacific is a large source because it is the site of large-scale upwelling, a process which brings cold water from depth to the surface. These waters are in turn heated by solar radiation, and because the solubility of CO_2 is strongly temperature dependent (CO_2 is less soluble in warm water), it is lost to the atmosphere. Conversely, polar waters are in general sinks for CO_2. Waters there lose heat to the atmosphere, and thus are able to absorb more CO_2. A topic of intense debate is the possible decrease in carbon flux to the waters of the Southern Ocean resulting from recent increases in wind strength, which may have altered the ocean's ability to remove CO_2 [5]. Such changes potentially would have profound impacts on the global carbon budget. At the present time the ocean is a net sink for atmospheric carbon dioxide, and has sequestered at least 25% of all anthropogenic emissions to date.

Ocean Acidification

Recently, great concern has been expressed about the increasing concentrations of CO_2 in the ocean, since its absorption decreases the pH, leading to ocean

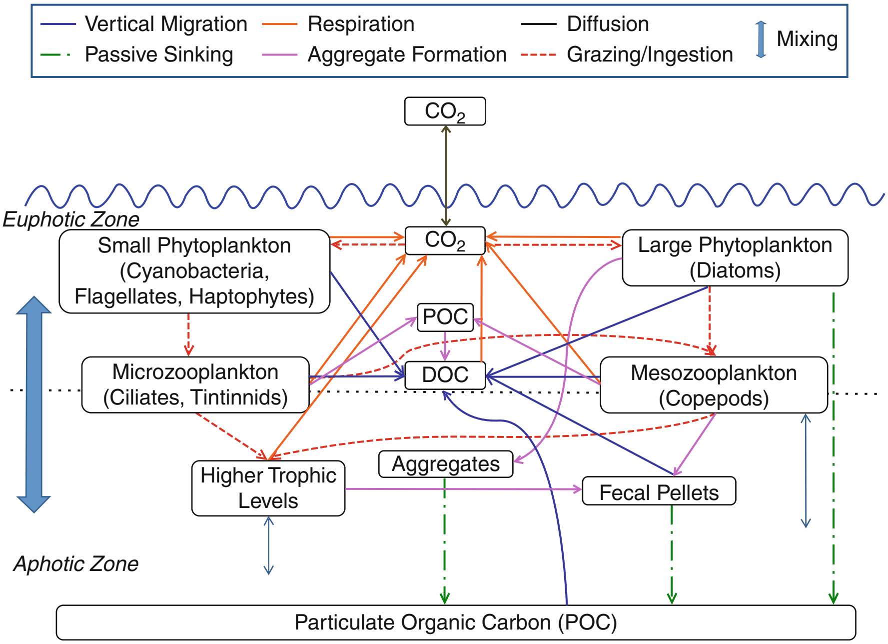


Fig. 12.2 Schematic of the biological pump showing the biological and chemical components and processes involved in the transformation of carbon dioxide (CO_2) to organic matter, and the subsequent transformation, movement, and oxidation of particulate organic carbon (POC) and dissolved organic carbon (DOC). The CO_2 is absorbed from the atmosphere across the air-ocean interface (*wavy lines*) and is transformed by processes in the euphotic (above *dashed line*) and aphotic zone (below *dashed line*). The migration of zooplankton and higher trophic levels within the water column (*light blue lines*) and unidirectional passive sinking of particles of different sizes to depth (*green dot-dashed line*) redistribute organic material. Processes of grazing/ingestion (*red dashed line*), aggregate formation (*red line*), respiration and CO_2 generation (*orange line*), physical mixing (heavy *blue line*), and solubilization, and DOC generation (*dark blue line*) modify the rate at which POC is exported to depth from the surface waters. The POC pool at depth is generally composed of unidentifiable, small particles, whereas the POC pool in the surface is composed of recognizable biota (bacteria, phytoplankton, zooplankton) and variable amounts of detritus

acidification [5, 6]. A decrease in pH would seriously impact calcification, likely increasing dissolution of $CaCO_3$ found in skeletons and shells because the material is unprotected from seawater, and decreasing the rate at which calcification can occur by altering the concentrations of the necessary minerals in the water column. As a result, decreased pH has a great capacity to alter the ecology of marine systems such as coral reefs. In addition, decreased pH levels have been shown to alter the growth, reproduction, efficiency, and survival of those organisms that require $CaCO_3$ to survive, and these effects vary among organisms, suggesting that substantial and unexpected impacts on biodiversity could occur [7].

It is now recognized that many phytoplankton can remove only CO_2 for use in photosynthesis. Under preindustrial pH levels, free CO_2 levels could have been at limiting levels, particularly for conditions that produced high concentrations of algae, because photosynthesis naturally increases the pH level. Decreased pH and increased absolute CO_2 levels arising from current conditions might reduce this limitation. Because there is substantial variability among species of phytoplankton in their response to increased CO_2, planktonic biodiversity is at risk [8]. However, certain algal functional groups, such as nitrogen-fixing cyanobacteria, positively respond to increased CO_2 concentrations by increasing their growth and photosynthesis, whereas others can not. Similarly, at least one species of toxin-producing dinoflagellate demonstrated increased growth and modified elemental ratios under increased CO_2 conditions [9], suggesting the possibility of an enhancement of occurrences of harmful algal blooms in the future. Because the marine carbon cycle is intimately linked with the biogeochemical cycles of nitrogen, phosphorus, silicon, and iron, these interactive effects make it extremely difficult to predict what future decreases in oceanic pH will generate. Oceanographers have recognized that increased inorganic carbon levels can have subtle effects on the biota, and much work is being done to document and quantify these effects.

The Biogeochemical Cycles of Nitrogen and Phosphorus

Although the early work of Redfield [1] clearly differentiated between the sources of nitrogen and phosphorus and the regulation of their turnover, they are linked in nature by the processes operating in the biological pump (Fig. 12.2). Despite this coupling, as well as their linkage to carbon, there are a number of features that distinguish them.

Nitrogen occurs in three reactive, inorganic forms in the ocean: nitrate (NO_3^-), nitrite (NO_2^-), and ammonium (NH_4^+) and the processes that transform and modify these forms make up the nitrogen cycle (Fig. 12.3). The nitrogen cycle has five major pathways that result in changes in the availability of nitrogen that can be used by plants. Nitrogen fixation removes gaseous nitrogen from the atmosphere, which is then converted by a series of reactions to forms that can be used for plant growth. In the ocean this process occurs primarily in tropical and semitropical environments, and the major algal species responsible for this transformation is *Trichodesmium*. Denitrification results in the reduction of NO_3^- to gaseous nitrogen, usually mediated by bacteria, and results in the loss of nitrogen available for phytoplankton in oceanic systems. These two processes are the primary means by which the ocean biota controls nitrogen biogeochemistry.

Nitrogen assimilation is the process by which nitrate (NO_3^-) and ammonium (NH_4^+) are removed from the water by phytoplankton. Ammonium is energetically favored for uptake because it does not have to be reduced intracellularly, but nitrate often occurs in greater concentrations, particularly in areas of upwelling or deep vertical mixing. Ammonium inhibits nitrate uptake, but the degree of inhibition varies with the relative concentration of the two nutrients. Ammonification generates NH_4^+ by the cleaving of amine groups from organic nitrogen. Because many marine organisms excrete ammonium, the vertical distribution of NH_4^+ can

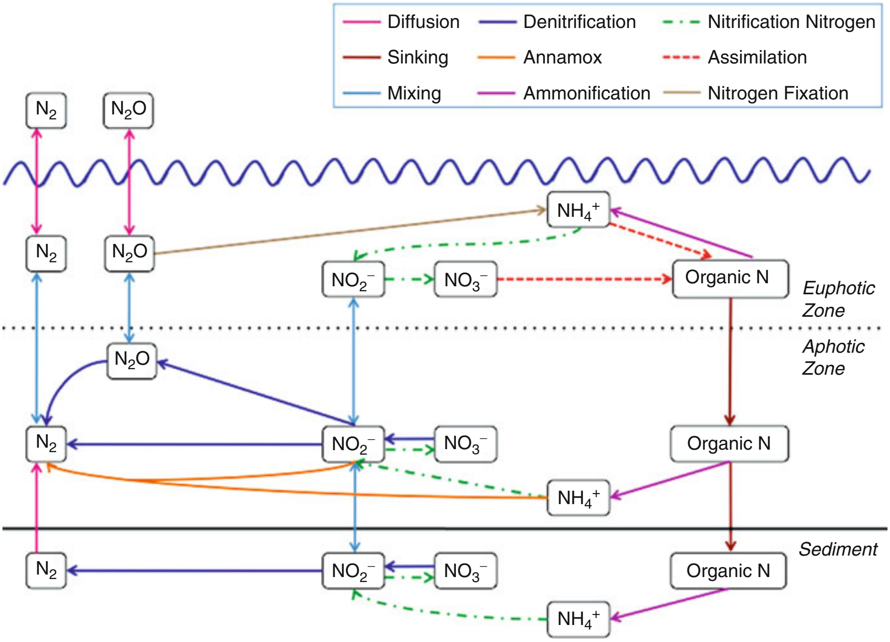


Fig. 12.3 Schematic of the nitrogen cycle showing the nine key mechanisms by which nitrogen moves through the water column, which are denitrification (*dark blue line*), nitrification (*green dash-dot line*), nitrogen assimilation (*red dashed line*), nitrogen fixation (*brown line*), ammonification (*purple line*), annamox (*orange line*), mixing (*light blue line*), diffusion (*pink line*), and sinking (maroon line)

depend on the distribution of heterotrophs, such as copepods, which is variable. Nitrification is the production of NO_3^- from ammonium. Earlier work suggested that this was a relatively slow process, but more recent investigations suggest that the oxidation of NH_4^+ and production of nitrate is quite rapid, particularly in tropical waters.

The different transformations result from different organisms and some require specific types of environmental conditions (Table 12.1). Denitrification and nitrogen fixation are anaerobic processes, which occur only in the absence of oxygen. Oceanic systems, ranging from estuarine to open ocean, provide sites for denitrification and as a result are depleted of oxygen. These oxygen-minimum regions are characterized by large vertical fluxes of organic matter, which heterotrophic bacteria oxidize and release nitrogen, consuming the available oxygen in the process. Anoxic conditions also occur in sediments where oxygen is depleted by aerobic metabolism. An unusual biological adaptation allows for nitrogen fixation (an anaerobic process) to occur in surface waters with high levels of oxygen. Some organisms (e.g., the cyanobacterium *Trichodesmium)* form extensive patches or tufts. These tufts, by virtue of their own metabolism, unusually thick cell walls and biochemical modifications of specialized cells where N_2 fixation occurs, create

Table 12.1 Summary of the major processes in the nitrogen cycle, the organisms responsible for the different processes, and the environmental conditions necessary for each process

Process	Organism(s) responsible	Necessary environmental condition
Nitrogen fixation $[N_2(\text{gas}) \rightarrow \text{reduced N}]$	Cyanobacteria, nitrogen-fixing bacteria	Absence of O_2; light for cyanobacteria
Denitrification $[NO_3^- \rightarrow N_2(\text{gas})]$	Denitrifying bacteria	Absence of O_2
Ammonification $[N_{\text{org}} \rightarrow NH_4^+]$	Heterotrophic organisms	Presence of O_2
Nitrogen assimilation $[NO_3^- \rightarrow N_{\text{org}}; NH_4^+ \rightarrow N_{\text{org}}]$	Large phytoplankton/diatoms for NO_3^- uptake; small phytoplankton for NH_4^+ uptake	Light
Nitrification $[NH_4^+ \rightarrow NO_3^-]$	Bacteria	Presence of O_2
Annamox $[NH_4^+, NO_2^- \rightarrow N_2(gas)]$	Bacteria	Absence of O_2

a microzone of very low oxygen, thus allowing nitrogen fixation to proceed. Other, smaller cyanobacteria have unusual biochemical adaptations that allow them to fix N_2 as well, despite living in oxygen-saturated water.

Recently a new nitrogen transformation, the annamox pathway, has been described in which anaerobic bacteria oxidize ammonium and nitrite directly to gaseous nitrogen, providing a second means by which nitrogen is “lost” from the nitrogen cycle [10]. This pathway has been found to be quantitatively important in regions such as the Peruvian and Arabian Sea oxygen-minimum zones [11, 12]. Because 30–50% of global nitrogen “losses” occur in these types of regions, elucidation of this process, its oceanographic controls, and the absolute rates, have important implications for the global nitrogen cycle.

The Biogeochemical Cycle of Iron

The understanding of the role of iron in the ocean has undergone a dramatic revision in the past few decades. Until recently data on absolute iron concentrations were seriously compromised by the difficulty of obtaining samples without contamination. As the collection and sampling aspects were greatly improved, the ability to quantify concentrations of iron in the oxygenated waters of the ocean decreased dramatically. Coincident with increased realization and acceptance of the vanishingly low concentrations of iron was the hypothesis that iron could, and does, regulate phytoplankton growth and productivity over large areas of the ocean [13]. Indeed, the hypothesis appeared to explain a number of oceanic features that were only partially explained. For example, large areas of the ocean, such as the Southern Ocean, the equatorial Pacific, and the north Pacific, have substantial standing stocks of nitrate and phosphate, as well as adequate irradiance, but exhibit very low standing stocks of phytoplankton (high-nutrient, low-chlorophyll regions, or HNLCs). During glacial-interglacial periods, atmospheric concentrations of CO_2 showed substantial variations and were strongly negatively correlated with iron deposition [14]. Thus iron limitation could explain CO_2 variations over geological time as well. Given that iron is the fourth most abundant element on earth, how can such low concentrations exist in the ocean, and how did oceanographers unequivocally demonstrate the ecological importance of iron?

Iron is derived from terrestrial and hydrothermal sources, but upon entry into oxygenated, saline waters, it rapidly forms iron oxides. The precipitates are largely insoluble under aerobic conditions, and attach to particles or remain in the water as colloids. The colloids can be solubilized by irradiance, contributing to a pool of dissolved inorganic iron, which consists of two forms, Fe^{+2} and Fe^{+3}. Both of these ions can be removed by plankton for their growth, although Fe^{+2} is generally oxidized to Fe^{+3} and kept at low levels. The mean ocean concentration of dissolved inorganic iron in the upper 200 m of the ocean is 0.07 nmol kg^{-1} [12]. Both forms can also be chelated by organic molecules, and thus become part of the dissolved ferro-organic pool. In general, there are two classes of organic ligands that bind with iron, a strong-binding ligand and a weak-binding ligand. The latter exchanges iron easily with biota, and thus makes iron bioavailable. There is also a class of special ligands called siderophores, which are low molecular weight organics that are produced and excreted primarily by prokaryotic organisms (bacteria, cyanobacteria) and that bind dissolved inorganic iron [15]. The ferro-ligand complex can be assimilated by bacteria, phytoplankton, and cyanobacteria, and the iron incorporated into a variety of cellular processes. Transformations among all of these pools are both biologically and irradiance mediated; entirely different transformations and equilibria are established in anoxic waters and sediments.

Iron in ocean surface waters derives from either atmospheric or deep ocean sources. Atmospheric deposition varies by latitude (proximity to terrestrial sources) and temporally (dependent on source region wind variability). Aerosols can be measured by satellite-borne sensors, which have shown that some oceanic systems receive substantial periodic depositions of iron from industrial sources (the North Atlantic) and from dust derived from terrestrial deserts in China (the western Pacific) and the Sahara in Africa (the coast of North Africa). Dissolution of aerosols in ocean water (fractional solubility) depends on the type of mineral in the aerosol, and can range from <1–90% [16, 17]. Small aerosol particles can rapidly aggregate with biological particles and exit the surface layer by sinking. Residence times for particulate iron can be as short as 6 days [17]. Conversely, other regions are rarely impacted by atmospheric deposition events (e.g., the Southern Ocean, the equatorial Pacific) by virtue of large-scale wind patterns that isolate them from terrestrial sources. These regions have their iron inputs driven by oceanographic processes such as deep vertical mixing and upwelling. Given the spatial and temporal variability in both of these processes, it is not surprising that surface water concentrations of iron are also highly variable.

Mesoscale Iron Fertilization Experiments

In the 1990s a series of large-scale ocean manipulations were undertaken to test the hypothesis that iron limited phytoplankton growth in the tropical Pacific. Two competing hypotheses were offered to explain the equilibrium concentrations of high concentrations of nitrate and low phytoplankton biomass which were (1) limiting levels of bioavailable iron and (2) rates of loss processes from grazing kept phytoplankton standing stocks at low levels. To test these, in situ additions of iron were planned for limited regions of the ocean. The passive tracer sulfur hexafluoride (SF_6), which can be detected at very low levels, was added with the iron so that the enriched water could be followed over time. The first iron enrichment experiment produced contradictory results. The photosynthetic capacity of phytoplankton showed a clear enhancement that was correlated with iron additions, but nitrate and CO_2 concentrations were unaffected [18, 19]. Further analysis showed that upon initial iron enrichment, the iron dropped to extremely low levels because colloid formation rapidly converted soluble iron to insoluble iron oxides, and the fertilized water patch was subducted to depth, which removed the iron-enriched water from the high irradiance euphotic zone required for nutrient assimilation. To further test the two hypotheses, the experiment was repeated, and this experiment clearly demonstrated the critical role of iron in limiting phytoplankton growth in high-nutrient, low-chlorophyll waters. Iron was added repeatedly to the patch of water at 3-day intervals for almost 2 weeks [20], and the response of the surface water was clear, showing decreased nitrate (which dropped to zero), decreased CO_2, increased phytoplankton biomass and photosynthetic activity, and a quantifiable decrease in iron concentrations. That is, the concentration and supply of iron was nevertheless the essential feature in driving the carbon and nitrogen cycles of the equatorial Pacific Ocean.

Subsequent similar iron enrichment experiments have been conducted in other HNLC regions in the Southern Ocean and the North Pacific. The former is extremely important to global biogeochemical cycles, as it is the site of deep and intermediate water mass formation, and thus regulates the concentrations of inorganic nutrients in much of the world's surface waters. As an example, models suggested that if all the inorganic nutrients were utilized (by iron fertilization) in the Southern Ocean that within 300 years the waters being upwelled in the eastern tropical Pacific would be greatly reduced in nutrient levels, and thus decrease productivity of commercially important higher trophic levels and marine mammals dependent on ecosystem processes in that region [21]. In all iron enrichment experiments to date, substantial and positive responses to additions of inorganic iron were observed, and while the details among experiments differ (and the causes debated), it is now accepted that iron plays a major role in the biogeochemistry of the ocean [22].

The Biogeochemical Cycle of Silicon

Silicon, despite being a nutrient for only one major functional group of phytoplankton (diatoms), is a major factor in regulating other biogeochemical cycles, such as carbon. This is because diatoms are extremely important primary producers, generating approximately as much oxygen on an annual basis as do pine trees in terrestrial systems. In addition, diatoms are among the largest forms of phytoplankton, and hence can sink passively to depth. Diatoms also produce transparent exopolymer particles, which serve as the primary mechanism for aggregating particles in the ocean's surface layer, thus producing large, rapidly sinking particles that are the major component of organic carbon and nitrogen flux to deeper water (Fig. 12.2). Finally, diatoms are also heavily grazed by herbivorous organisms, and serve as a means to transfer photosynthate to the large organism-based food web. All of these characteristics contribute to the substantial importance of diatoms in the ocean.

Silicon is a major component of rocks and terrestrial minerals, and as a result the inputs to the ocean in riverine waters are substantial. However, silicon is not readily dissolvable, and dissolved silicon, which occurs as $Si(OH)_4$, remains at relatively low levels. Aeolian and oceanic weathering of seafloor rocks also constitutes a significant source of dissolved silicon. Silicon also is found in high concentrations in waters exiting hydrothermal vents, and while quantitative estimates are uncertain, the contribution of this source to total silicon inputs is likely to be significant.

Silicon is incorporated into diatoms and other marine organisms as opal $(Si(OH)_4nH_2O)$, which is slightly more soluble than pure SiO_2 and undersaturated in all ocean waters. Opal is found in the sediments as siliceous deposits of biogenic origin; these deposits are largely focused in the Southern Ocean's polar front region [23]. Silicon is recycled within the water column, but rates of this cycling are modest, and silicon regeneration is often markedly uncoupled from that of carbon and nitrogen in some regions. The reason for this appears to result from the different controls of each: organic matter regeneration is largely biologically mediated (by heterotrophic processes), whereas silicon regeneration is regulated by temperature [24]. As a result, in polar regions a large fraction of the organic matter that sinks from the euphotic zone is regenerated in the upper 250 m, whereas a substantial amount of silicon sinks to a greater depth as biogenic particles. This uncoupling contributes to the formation of large zones of biogenic silica deposits in polar regions and are reflective of surface layer diatomaceous productivity. In a more recent reanalysis of the global silicon budget, it was concluded that the deposition of silicon in continental margins may have been greatly underestimated [25]. If this were true, then the coupling between the silicon and organic matter budgets would be even stronger than previously thought.

An additional mechanism to couple the biogeochemistry of silicon and organic carbon is the presence of an organic membrane that covers diatom frustules [26]. Silica dissolution does not begin until this membrane is degraded by bacteria, which decreases the time for the dissolution of opal during the transit of a particle through the water column (ca. 3,000 m). Sinking rates of large aggregates are ca. 200 m day^{-1}, so that a reduction in the already low rate of dissolution by the necessity for organic degradation can decrease dissolution of silica markedly. Similar effects of grazing can occur, as fecal pellets are usually composed of an organic pellicle that must be degraded prior to chemical silica dissolution.

As with other nutrients, silicic acid has substantial interactions with other elements, such as nitrate and iron. Under iron-limiting conditions, diatoms continue to assimilate silicon, but because iron is needed in the enzymes used for nitrate assimilation, nitrate uptake decreases [3]. As a result, Si:N ratios increase by nearly an order of magnitude in diatoms under iron limitation and elevated ratios observed in natural systems have been used to infer iron limitation.

The Biogeochemical Cycle of Sulfur

In marine systems sulfur is largely present in its most stable form, which is sulfate (SO_4^{-2}). Sulfate is present in high concentrations in most marine systems, and relatively low concentrations are required by organisms to survive [27]. As a result, sulfur does not normally become growth limiting. Sulfate concentrations in marine systems are primarily controlled by physical rather than chemical processes. Variations in concentration only have a significant biological impact in anoxic zones where sulfate reduction occurs [2]. Sulfur is also present as other inorganic (H_2S) and organic (dimethylsulfoniopropionate (DMSP), dimethylsulfide (DMS), carbonyl sulfide (COS), and methanethiol (MeSH)) forms. Sulfate is transformed into these compounds via the sulfur cycle, which operates primarily in the photic zone of the upper water column, in the sediments, and around hydrothermal vents (Fig. 12.4).

In aerobic environments sulfur is converted between inorganic compounds (sulfate and hydrogen sulfide) and organic sulfur compounds including DMSP, DMS, COS, and amino acids. Most algae and bacteria use sulfur assimilation to form amino acids, such as cysteine and methionine [27]. Some phytoplankton species, particularly prymnesiophytes and dinoflagellates, use methionine to produce DMSP, a compound with antioxidant properties [28, 29]. DMSP can be released into the water and subsequently used to produce amino acids through assimilation by bacteria or phytoplankton, including some species of diatoms and cyanobacteria, demethylated by bacteria to produce MeSH, or oxidized into DMS and acrylic acid [30]. DMS is either broken down in the water into sulfate through bacterial uptake or photooxidation, or is volatilized into the atmosphere, where it can act as an important aerosol [27].

The sulfur cycle in ocean sediments can be divided into reactions that occur in the upper oxic layer and those that occur in the lower, oxygen-depleted (anoxic) region. In the anoxic sediments, sulfur-reducing bacteria carry out anaerobic respiration using sulfate or sulfur-containing organic compounds to oxidize organic matter, resulting in the production of sulfide, typically as H_2S, a form of sulfur that

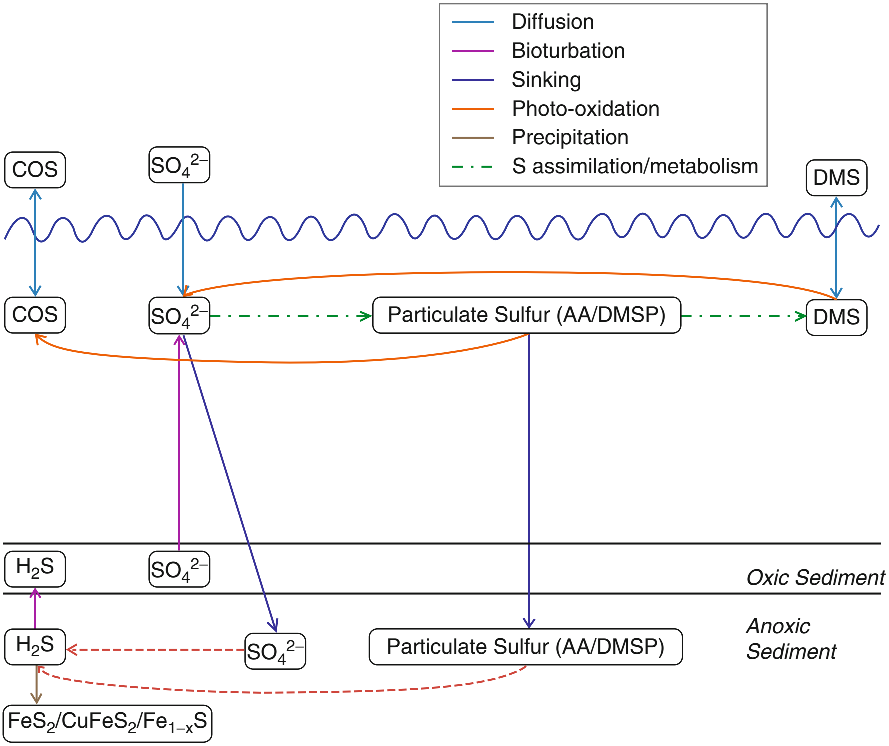


Fig. 12.4 The sulfur cycle involves transformation of sulfur in the water column through physical mechanisms such as diffusion (*light blue line*), bioturbation (*purple line*), and sinking (*dark blue line*); chemical mechanisms of photooxidation (*orange line*), and precipitation (*yellow line*); and biological mechanisms of sulfur assimilation and metabolism by phytoplankton (*green dash-dot line*) and reduction and oxidation by bacteria in the sediment (*orange dashed line*)

is highly toxic to most organisms. In the deeper layers of the sediment, sulfide reacts with iron and precipitates as iron sulfides such as pyrite (FeS_2) [30]. Some sulfide remains in the sediment, and, when mixed back into the oxic zone through processes such as bioturbation, is quickly oxidized by sulfur-oxidizing bacteria into sulfate, which can then remain in the sediment or be released into the overlying water [31]. Sulfur oxidation and reduction by bacteria in the sediment are also important to the functioning of the nitrogen cycle in oxygen-minimum zones [32]. In these environments, sulfate reduction provides a significant amount of the ammonium used in the anammox reaction in anaerobic environments, and nitrate reduction may be coupled to sulfide oxidation, indicating that the anaerobic mechanisms in the sulfur cycle may also be important in the nitrogen cycle [32].

The presence of hydrogen sulfide around hydrothermal vents has resulted in the development of unique organisms with the ability to use the energy contained in hydrothermal fluids to produce organic compounds through chemoautolithotrophy [33]. At hydrothermal vents seawater comes into contact with magma from the earth's interior, which cools and forms reduced sulfur compounds [2]. The sulfate in seawater then reacts to form hydrogen sulfide as well as sulfur-containing minerals such as pyrite (FeS_2), chalcopyrite ($CuFeS_2$), and pyrrhotite ($Fe_{1-x}S$) [2], which form the surface chimney structure that is characteristic of hydrothermal vents. The hydrogen sulfide provides the energy, rather than light, for the chemoautrophic microorganisms that form the base of the hydrothermal vent food web [33]. Some species of microorganisms can operate in aerobic conditions, using oxygen as the electron acceptor, while others have the ability to carry out this reaction in anaerobic conditions, using nitrate, sulfate, or sulfur as the electron acceptor [2]. These organisms survive in symbiotic relationships with other organisms living near the hydrothermal vents. The microorganisms, which are endemic to hydrothermal vent environments, allow unique communities to develop and help maintain the oceanic sulfur cycle by transforming hydrogen sulfide released by the hydrothermal vents into sulfate [33].

The Biogeochemical Cycle of Oxygen

Oxygen is involved in all nutrient cycles, and its presence or absence dictates the reactions that will occur in a specific marine environment. Oxygen gas can be introduced into marine environments across the air-sea interface (e.g., by diffusion). However, oxygen concentration is controlled by the biological processes of photosynthesis and respiration, and by physical processes such as mixing within the water column. In the euphotic zone, phytoplankton photosynthesis produces oxygen, which is then used as the electron acceptor to conduct aerobic respiration. This process is carried out by both autotrophic and heterotrophic organisms throughout the water column.

Oxygen concentration generally decreases with depth in the ocean. Photosynthesis can only be carried out in the lighted parts of the water column, but respiration continues throughout the water column. As the organic matter from the surface layers sinks, it is taken up by organisms and used to conduct respiration, depleting oxygen levels. Some marine environments, particularly in marine sediments, are suboxic, with oxygen concentrations less than 0.2 ppm (but still detectable), or anoxic, with oxygen concentrations below detectable levels [2]. Organisms survive in these environments by using anaerobic respiration, in which compounds such a nitrate, sulfate, iron, or even organic matter are used as alternative electron acceptors to oxygen [2].

Anoxic zones are not limited to marine sediments, with increasing attention being paid to decreasing oxygen concentrations in previously oxygen-rich areas of the ocean. Hypoxic zones, marine environments with oxygen concentrations below 2 mg L^{-1}, typically form when primary productivity is high, leading to increased organic matter in the system and increased respiration, and when mixing throughout the water column is low, preventing the oxygen in the upper water column from reaching lower layers [34–36]. Hypoxic zones have been increasing in frequency, including the Gulf of Mexico and Chesapeake Bay [34, 35]. Factors such as eutrophication due to increased fertilizer or wastewater runoff have lead to the development of hypoxic conditions in systems already susceptible due to vertical stratification of the water column [34]. Thus, the disruption in the typical oxygen cycle and the lack of an anaerobic respiration mechanism in most marine organisms can result in serious consequences for the composition and productivity of the marine food web community in these hypoxic zones.

Future Directions

Studies of Biogeochemical Cycles

In the past two decades, a number of large, interdisciplinary programs were conducted to obtain biogeochemical data on appropriate time and space scales so that mathematical models of global climate change can accurately represent the complex processes of elemental cycles. One such program, the Joint Global Ocean Flux Study (JGOFS), which occurred from 1987 to 2003, was international in scope, and undertook coordinated, multidisciplinary, international studies in the equatorial Pacific, the north Atlantic, the Arabian Sea, and the Southern Ocean, and coordinated multidisciplinary national programs in a range of coastal and open ocean environments. The JGOFS project was designed to assess the carbon cycle, but because all elemental cycles are closely linked, insights were gained into the understanding of nitrogen, silicon, and iron cycles as well. The JGOFS program also had a significant synthesis and modeling component that was intended to integrate the data sets from the multidisciplinary studies and to develop mathematical models of increased complexity and biological realism. In addition to providing a wealth of publicly available data, the JGOFS program served as a model for large, multidisciplinary studies of ocean processes.

The results and understanding from the JGOFS program provided the basis for the Integrated Marine Biogeochemistry and Ecosystem Research (IMBER) Project, which was initiated in 2001 by the International Geosphere-Biosphere Program and the Scientific Committee on Oceanic Research. The science goals of the IMBER project extend the investigation of marine biogeochemical cycles to include the influence of feedbacks with marine food webs and the consequences for marine ecosystems. Central to the IMBER goal is the development of a predictive understanding of how marine biogeochemical cycles and ecosystems respond to complex forcings, such as large-scale climatic variations, changing physical dynamics, carbon cycle chemistry and nutrient fluxes, and the impacts of marine harvesting. IMBER science is making new advances in understanding marine systems by bringing together the natural and social science communities to study key impacts and feedbacks between the marine and human systems. The emerging recognition of human interactions as integral parts of marine ecosystems is providing the direction for future integrative research designed to understand and sustain ocean systems as environmental change and its associated uncertainties occur.

Role of Modeling

Mathematical models provide an approach for integrating and synthesizing the knowledge and understanding obtained from measurements of oceanic biogeochemical processes. The use of biogeochemical models in ocean research has a long history [37, 38] but their use was advanced significantly in the early 1990s when a model that simulated nitrogen cycling through the lower trophic levels in the oceanic mixed layer became generally available [39], which subsequently has provided the basis for the coupled circulation-biogeochemical models that are now embedded in regional, basin, and global scale models.

The skill of the current generation of biogeochemical models is sufficient to allow projections of future states that may result from climate variability and the oceanic uptake of anthropogenic carbon [40–42]. The patterns and distributions emerging from these simulations show shifts in phytoplankton distributions and marine biomes, alteration of phytoplankton species assemblages, and modified lower trophic level community structure [43–45], all of which have direct and important consequences for biogeochemical cycling. Simulations of the effect of increasing atmospheric CO_2 and its uptake by the ocean show reductions in ocean pH and in saturation levels of calcium carbonate, which have serious consequences for many marine organisms [46].

Advances in conceptual understanding, modeling techniques, and data availability have made predictive marine biogeochemical models a feasible goal [47]. However, modeling for prediction is still rapidly developing and much remains to be done in generating appropriate frameworks and in collection of data sets that support predictive modeling for marine biogeochemical cycling [48, 49].

Bibliography

Primary Literature

1. Redfield AC (1958) The biological control of chemical factors in the environment. Am Sci 64:205–221
2. Libes SM (2009) Introduction to marine biogeochemistry. Academic, Amsterdam
3. Hutchins DA, Bruland KW (1998) Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. Nature 393:561–564
4. Sunda WG, Huntsman S (1997) Interrelated influence of iron, light, and cell size on growth of marine phytoplankton. Nature 390:389–392
5. Le Quéré C et al (2007) Saturation of the Southern Ocean CO_2 sink due to recent climate change. Science 316:1735–1738
6. Doney SC, Fabry VF, Feely RA, Kleypas JA (2009) Ocean acidification: the other CO_2 problem. Ann Rev Mar Sci 1:169–192
7. Ries BR, Cohen AL, McCorkle DC (2009) Marine calcifiers exhibit mixed responses to CO_2-induced ocean acidification. Geology 37:131–134
8. Hutchins DA, Mulholland MR, Fu F (2009) Nutrient cycles and marine microbes in a CO_2-enriched ocean. Oceanogr 22:128–145

9. Fu F-X, Zhang Y, Warner ME, Feng Y, Hutchins DA (2008) A comparison of future increased CO_2 and temperature effects on sympatric *Heterosigma akashiwo* and *Prorocentrum minimum*. Harm Algae 7:76–90
10. Devol AH (2003) Solution to a marine mystery. Nature 422:575–576
11. Lam P et al (2009) Revising the nitrogen cycle in the Peruvian oxygen minimum zone. Proc Nat Aca Sci 106:4752–4757
12. Ward B et al (2009) Denitrification as the dominant nitrogen loss process in the Arabian Sea. Nature 451:78–81
13. Martin JH, Fitzwater S (1988) Iron deficiency limits phytoplankton growth in the northeast Pacific subarctic. Nature 331:341–343
14. Ducklow HW, Oliver JL, Smith WO (2007) The role of iron as a limiting nutrient for marine plankton processes. In: Melillo JM, Field CB, Moldan B (eds) Interactions of the major biogeochemical cycles. Island Press, Washington, DC, pp 295–310
15. Vraspir JM, Butler A (2009) Chemistry of marine ligands and siderophores. Ann Rev Mar Sci 1:43–63
16. Johnson KS, Gordon RM, Coale KH (1997) What controls dissolved iron concentrations in the world ocean? Mar Chem 57:137–161
17. Baker AR, Croot PR (2010) Atmospheric and marine controls on aerosol iron solubility in seawater. Mar Chem 120:4–13
18. Martin JH et al (1994) Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. Nature 371:123–129
19. Watson AJ et al (1994) Minimal effect of iron fertilization on sea-surface carbon dioxide concentrations. Nature 371:143–145
20. Coale KH et al (1996) A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific. Nature 383:495–501
21. Sarmiento JL, Gruber N, Brzezinski MA, Dunne JP (2004) High-latitude controls of thermocline nutrients and low latitude biological productivity. Nature 427:56–60
22. Boyd PW et al (2007) Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. Science 315:612–617
23. Treguer P, Nelson DM, van Bennekom AJT, DeMaster DJ, Lynaert A, Queguiner B (1995) The silica balance in the world ocean: a reestimate. Science 268:375–379
24. Nelson DM et al (1995) Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentations. Global Biogeochem Cycles 9:359–372
25. DeMaster DJ (2002) The accumulation and cycling of biogenic silica in the Southern Ocean: revisiting the marine silica budget. Deep Sea Res II 49:3155–3167
26. Biddle KD, Azam F (1999) Accelerated dissolution of diatom silica by natural bacterial assemblages. Nature 397:508–512
27. Sievert SM, Kiene RP, Schulz-Vogt HN (2007) The sulfur cycle. Oceanography 20:117–123
28. Andreae MO (1990) Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. Mar Chem 30:1–29
29. Sunda W, Kieber DJ, Kiene RP, Huntsman S (2002) An antioxidant function for DMSP and DMS in marine algae. Nature 418:317–320
30. Jorgensen BB (1977) The sulfur cycle of a coastal marine sediment (Limfjorden, Denmark). Limnol Oceanogr 22:814–832
31. Canfield DE, Farquhar J (2009) Animal evolution, bioturbation, and the sulfate concentration of the oceans. Proc Nat Acad Sci 106:8123–8127
32. Canfield DE, Stewart FJ, Thamdrup B, Brabandere LD, Dalsgaard T, Delong EF, Revsbech NP, Ulloa O (2010) A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. Science 330:1375–1378
33. van Dover CL (2000) The ecology of deep-sea hydrothermal vents. Princeton University Press, Princeton

34. Diaz RJ, Rosenberg R (2008) Spreading dead zones and consequences for marine ecosystems. Science 321:926–929
35. Scavia D, Justic D, Bierman VJ (2004) Reducing hypoxia in the Gulf of Mexico: advice from three models. Estuaries 27:419–425
36. Lam P, Kuypers MMM (2011) Microbial nitrogen cycling processes in oxygen minimum zones. Ann Rev Mar Sci 3:317–345
37. Riley GA (1946) Factors controlling phytoplankton populations on Georges Bank. J Mar Res 6:54–73
38. Riley GA (1947) A theoretical analysis of the zooplankton populations of Georges Bank. J Mar Res 6:104–113
39. Fasham MJR, Ducklow HW, McKelvie SM (1990) A nitrogen-based model of plankton dynamics in the oceanic mixed layer. J Mar Res 48:591–639
40. Orr JC et al (2001) Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models. Global Biogeochem Cycles 15:43–60
41. Doney SC et al (2004) Evaluating global ocean carbon models: the importance of realistic physics. Global Biogeochem Cycles 18:GB3017. doi:10.1029/2003GB002150
42. Boyd PW, Doney SC (2001) Modeling regional response by marine pelagic ecosystems to global climate change. Geophys Res Lett 29. doi:10.10292001GL014130
43. Follows MJ, Dutkiewicz S, Grant S, Chisholm SW (2007) Emergent biogeography of microbial communities in a model ocean. Science 315:1843–1846
44. Le Quéré C et al (2005) Ecosystem dynamics based on plankton function types for global ocean biogeochemistry models. Global Change Biol 11:2016–2040. doi:10.111/j.1365-2486.2005.1004.x
45. Sinha B, Buitenhuis ET, Le Quéré C, Anderson TR (2010) Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models. Prog Oceanogr 84:204–224
46. Orr JC et al (2005) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437:681–686
47. Friedrichs MAM et al (2007) Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. J Geophys Res 112:C08001. doi:10.1029/2006JC003852
48. Anderson TR (2009) Progress in marine ecosystem modeling and the "unreasonable effectiveness of mathematics." J Mar Sys 81:4–11
49. Ducklow HW, Doney SC, Steinberg DK (2009) Contributions of long-term research and time-series observation to marine ecology and biogeochemistry. Ann Rev Mar Sci 1:279–302

Books and Reviews

Baliño BM, Fasham MJR, Bowles MC (2001) Ocean biogeochemistry and global change. IGBP Science no 2, Stockholm (A popular summary of JGOFS main achievements)

Benitez-Nelson CR (2000) The biogeochemical cycling of phosphorus in marine systems. Earth-Sci Rev 51:109–135

Brandes JA, Devol AH, Deutsch C (2007) New developments in the marine nitrogen cycle. Chem Rev 107:577–589

Broecker WS (2009) Wally's quest to understand the ocean's $CaCO_3$ cycle. Ann Rev Mar Sci 1:1–18

Deutsch C, Brix H, Ito T, Frenzel H, Thompson L (2011) Climate-forced variability of ocean hypoxia. Science 333:336–339

Diaz RJ (2000) Overview of hypoxia around the world. J Environ Quality 30:275–281

Doney SC et al (2001) Marine biogeochemical modeling: recent advances and future challenges. Oceanography 14:93–108

Fasham MJR (ed) (2003) The role of the ocean carbon cycle in global change. Springer, Heidelberg, p 297

Gruber N, Galloway JN (2008) An earth-system perspective of the global nitrogen cycle. Nature 451:293–296

Hanson RB, Ducklow HW, Field JG (eds) (2000) The changing ocean carbon cycle: a midterm synthesis of the joint global ocean flux study, IGBP book series no 5. Cambridge University Press, Cambridge, 520 pp

IMBER Science Plan and Implementation Strategy (2005) IGBP Report N°52. IGBP Secretariat, Stockholm, 76 pp

Jickells TD et al (2005) Global iron connections between desert dust, ocean biogeochemistry, and climate. Science 308:67–71

Keeling RF, Kortzinger A, Gruber N (2010) Oxygen deoxygenation in a warming world. Ann Rev Mar Sci 2:199–229

Paulmier A, Ruiz-Pino D (2009) Oxygen minimum zones (OMZs) in the modern ocean. Prog Oceanogr 80:113–128

Ragueneau O et al (2000) A review of the Si cycle in the modern ocean: recent progress and missing gaps in the application of biogenic opal as a paleoproductivity proxy. Global Planetary Change 26:317–365

Sabine CL, Tanhua T (2010) Estimation of anthropogenic CO_2 inventories in the ocean. Ann Rev Mar Sci 2:175–198

Sunda WG (2010) Iron and the carbon pump. Science 327:654–655

Chapter 13
Species Competition and Predation

Peter Chesson

Glossary

Apparent competition	The tendency for an increase in the density of a species to increase the impact of a natural enemy on that same species or other species.
Competition	The tendency for an increase in the density of a species to have a negative effect on the survival or reproduction of individuals of the same species or of other species by reducing resource abundance, reducing access to resources, or by direct harm of one individual organism on another associated with resource acquisition.
Density dependence	The tendency for an increase in the density of a species to have a negative effect on the survival or reproduction of individuals of the same or different species. As used in this essay, the species in question are in the same guild. Competition and apparent competition are special cases of density dependence.
Feedback loop	A chain of species interactions from one member of a guild, through other species, back to a species in that same guild. Feedback loops transmit density dependence.
Guild	A group of species potentially co-occurring in the same locality and having similar ecology in the sense of depending on the same or similar resources, often seeking

---

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

P. Chesson (✉)
Department of Ecology and Evolutionary Biology, The University of Arizona,
Tucson, AZ 85721, USA
e-mail: pchesson@u.arizona.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_13,

	those resources in similar ways, and susceptible to the same or similar natural enemies. The standard of similarity in this definition is not precise, and varies depending on the purpose of the investigator.
Natural enemy	An organism that benefits its own reproduction or survival by harming the individuals of a given species, commonly by feeding on them.
Niche overlap	For any pair of species, the degree to which density dependence through feedback loops is concentrated between species compared to within species. It is measured by the quantity ρ which varies between zero for no overlap (no interspecific density dependence) and 1 for complete overlap (interspecific density dependence is on average equal to intraspecific density dependence).
Predator	A species that gains food by killing and consuming individuals of the species in the ecological guild in question.
Species average fitness	For a given species in a guild, it is a numerical measure of how well that species is adapted to the environment with the property that it predicts which species would dominate if the niche overlaps, ρ, were all equal to 1. It is normally related to the long-term average per capita growth rates of the species measured at fixed levels of competition and apparent competition. In this essay, the fitnesses κ are obtained from per capita growth rates at zero levels of competition and apparent competition, which are achieved by setting all members of a guild at zero density. These growth rates are then divided by scaling factors that correct for differences between species in their levels of sensitivity to competition and apparent competition.
Stable coexistence	The tendency of the members of a guild to recover when individually perturbed to low density, allowing their long-term persistence in the presence of interactions with other guild members.

Definition of the Subject

Competition and predation are key interactions between species, and are major foci of thought and study in community ecology. They are believed to be major forces structuring natural communities, having critical roles in the determination of species diversity and species composition, and are regarded as important drivers of evolutionary processes. The relationships between the niches of different species determine how they interact through competition and predation, which then have key roles in assembly of local communities, and their reassembly following perturbations. Niches define patterns of linkages between species, their resources, and their natural enemies. These patterns include how linkages change over time, and between different spatial locations, and define mechanisms by which similar species are able to coexist by their effects on competition and predation relationships. The human element in the environment has profound effects on these phenomena. Changing the environment shifts interactions between species, and profoundly modifies the structure of food webs. In the modern day, there is much community reassembly, potentially involving major shifts in competition and predation. Humans transport invasive species that act as predators, prey, and competitors with potentially major effects on the community reassembly process.

Introduction

Competition and predation are key species interactions that are believed to structure natural ecosystems and to have major roles in systems dominated by humans. Both of these interactions involve consumer–resource relationships in one form or another [1]. The relationship between a predator and one of its prey species is of necessity a consumer–resource relationship with the predator being the consumer and the prey its resource. Competition is mostly commonly resource competition, where several consumer species share one or more resources and compete for these resources [2–4]. A resource may or may not be a biological species. When the resources are biological species that are killed by the consumer, competition necessary involves predator–prey relationships. However, resource species may instead be grazed or browsed, parasitized or infected. In these cases, the consumer species are natural enemies of their resources, harming them without necessarily killing them, which generalizes the idea of a predator–prey relationship to a species–enemy relationship.

The resources of plants are generally not biological species, and indeed this is the case with the resources of most plants, which are instead broadly light energy, water, and chemical elements [5]. Plants are often thought of as requiring space to grow as a resource [4]. Space then provides their other needs. In general, not all space is equal, and plant species tend to be somewhat specialized, leading to the concept of safe sites [6] (places that satisfy the requirements for establishment, growth, and reproduction for a particular species), and the regeneration niche [5] (an elaboration of the idea of a safe site with a particular view to how species compete with one another). Sedentary animal species that either settle in a particular place and do not move, or establish territories, can also be regarded as having space as a resource [7]. Animal species require particular places for particular uses, such as nest holes and wallows, and use various dead organic and inorganic materials in their lives. These all count as resources if they are used or occupied by an individual to its benefit.

Both competition and predation are assumed to involve harm. In the case of predation, of course the predators benefit from the relationship, and prey are harmed because individual prey are killed. This does not, however, preclude some positive benefits of predation, or more generally of natural enemies, on species that are attacked, as these species may well be adapted to their natural enemies, and suffer some negative effects when natural enemies are removed. For example, grasses may be adapted to grazers that remove old, less productive biomass as well as enhancing nutrient recycling [8]. In the case of competition, harm is mutually negative for the participant species, though it is often lopsided with some species being harmed much more than others.

Resource consumption may reduce resource availability to individuals in species that depend on the resource. In this case, harm occurs when lowered resource availability leads to reduced fitness of individuals because they suffer directly by consuming less resource, have to expend more energy or materials to obtain the resource, take greater risks to obtain it, or have to divert time from other beneficial activities to do so. Competition is also assumed to occur by direct negative interactions between individuals seeking the resource. It is not necessary in such instances for resource consumption to lower resource availability, but the presence of other individuals decreases the ability of a given individual to consume resources, or in the process of seeking resources, individuals harm each other in other ways, for instance through fighting [9, 10].

Predation and species–enemy relationships necessarily have strong effects in ecosystems, as they form the paths of energy and material flows [1]. The role of competition between species is less obvious, often indirect, and frequently controversial [11]. Although it is easy to verify that a predator consumes a prey species, it is much more difficult to demonstrate that one consumer species harms another consumer species through their resource consumption activities. While numerous rigorous experimental studies have firmly established that competition between species is frequently a strong force in nature, the effects of interspecific competition on various community properties have been difficult to establish. An abundance of theoretical work provides hypotheses, but rigorously testing of them in nature has proved to be difficult and often controversial [12–14]. Thus, although competition as a strong force is well established, the effects of that force are not.

Several difficulties arise in the study of competition. First, the consequences of competition between species (interspecific competition) do not rest with its absolute strength but with its strength relative to competition within species (intraspecific competition) [15]. Competition within species constrains the tendency of one species to harm another. Second, competition can be constrained by other interactions, such as predation, in some cases limiting its effects, but potentially interacting with competition in complex ways [16, 17]. Third, the natural world is extremely variable in time and space. This variability not only makes clear trends difficult to discern, it potentially interacts with competition modifying the outcome [15]. Thus, although competition can be shown to be present and strong, and is believed to have important implications for numerous community phenomena, clear tests of predictions have often been elusive.

The predictions from predation are most often of a different character from those of competition, yet they need not be. Competition comes from the interactions between consumers as a result of consumption of shared resources. Turning that on its head, when the resources are prey, they affect each other indirectly by providing food for common predators, potential increasing danger from those predator species. This indirect interaction between prey species is known to have analogous effects to competition, and has thus been termed apparent competition [18]. Although first suggested in a single-species context many years ago [19], it has only recently been generally understood, and so has received far less direct study than competition, but it is also likely that studies of competition have inadvertently included the effects of apparent competition: protocols to cleanly separate the two effects have yet to be developed.

Conceptual difficulties have plagued understanding of how competition and predation structure communities owing to the intrinsic complexity of the subject. The theory, however, has been making steady progress, and a much more comprehensive theoretical framework is available now than even a few years ago. The body of this essay explains the fundamentals of how competition and predation are hypothesized to structure communities in the light of these recent advances. Included are the intricacies introduced by the complex behavior organisms, and the often conceptually difficult area of how predictions about competition and predation can be made in the face of temporal and spatial variation in the physical environment, and in the presence of fluctuations in the populations of the organisms themselves. Applications to a theory of invasive species are then presented. Invasions of alien species represent perturbations to natural systems that can lead to a process of reassembly of communities of organisms. Competition and predation are believed to have large roles in this reassembly process, and explaining it is a critical challenge in community ecology. Because of the impacts of alien species on native communities, there are major implications for conservation biology also. Finally, challenges in the study of competition and predation, and promising future directions, are presented.

Feedback Loops

The fundamentals of competition and predation can be best understood in terms of feedback loops within a food web. Figure 13.1 shows a simplified food web, which should be considered as part of a food web rather than being any reasonably complete web that one might find in nature. It shows three trophic levels allowing understanding of how the middle trophic level is affected by the trophic levels above and below, which represent, respectively, predators and resources of that middle trophic level. Species in the same trophic level, by sharing trophic position, have strong similarities in their ecology. There may be other ecological similarities, and also differences within a trophic level, in the way the species relate to other trophic levels and to other elements in their environment. Species with similar ecology are commonly referred to as a *guild* and here the middle trophic level is the focus of discussion and is referred to as the focal guild, or just “the guild.”

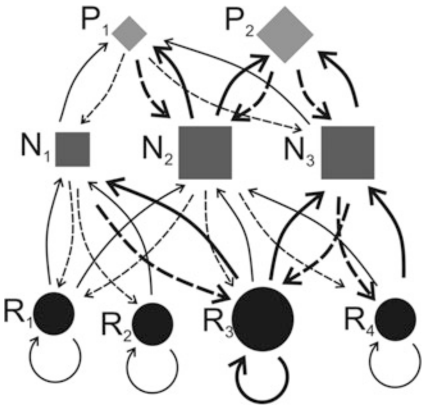


Fig. 13.1 Representation of a subset of a food web with arrows showing directions of effects from one species to another. Differences in the thicknesses of arrows are meant to indicate differences in the strengths of effects. The different symbols distinguish trophic levels, and their size differences are meant to imply differences between species in population size within a trophic level. The letters *R* and *P* indicate that they are respectively resources and predators of the species, *N*, in the middle trophic level, with subscripts labeling species within a trophic level. These subscripted letters serve simultaneously as species labels and as the population densities of the species. The recursive arrows for resources mean they experience direct intraspecific competition of some form giving density feedback to themselves, but not direct interspecific feedback within the resource trophic level. An assumption like this is common in the Lotka–Volterra models discussed in this essay, and may also be applied to the predators too

How species in a guild interact with each other, and ultimately coexist or instead exclude each other, leads to an understanding of how communities are structured, that is, an understanding of the relationships between the traits of different species that allow them to come together to form a long-lasting community, or in other words are assembled [20, 21]. Competition can be understood by considering the linkages between focal-guild species and their resources in the trophic level below. Linkages with the tropic level above lead to an understanding of apparent competition [22].

The arrows in the diagram show directions of effects. A species benefits from arrows pointing to it from a lower trophic level and suffers from arrows pointing to it from a higher trophic level. Chains of arrows define pathways of effects. Competition and apparent competition can be understood by pathways leading from the middle trophic level back to that level. For example, the species N_2 has pathways going from itself to each resource species, R, back to itself. These pathways contribute to intraspecific competition for N_2. The idea is that an increase in the density of N_2 leads to greater consumption of each resource species, R, reducing their densities and thereby reducing the availability of these resources for N_2. Pathways from these resources back to other species contribute interspecific competition. Thus, the pathways from N_2 through R_3 and R_4 back to N_3 lead to interspecific competition for N_2 on N_3. Through these various pathways, increasing the density of N_2 feeds back negatively to itself and to other species in the same trophic level. Critical to understanding of the ultimate effects of competition is the extent to which negative feedback of a species on itself limits its ability to increase and harm other species.

Apparent competition is understood by entirely analogous effects. In this case, we note that N_2 is preyed upon by P_1 and P_2. An increase in N_2 benefits both P_1 and P_2. These predators may therefore increase in abundance, and as a consequence, predation on the second trophic level will increase. Again, increasing the density of a species in the middle trophic level feeds back negatively to itself and to other species. Thus, in a food web, multispecies density dependence occurs when changes in the density of a species within a given trophic level feed back to that trophic level through linkages to other trophic levels, modifying the growth rates of the species. The term "density dependence" is traditionally used in ecology just for feedback from a species to itself, but understanding of the joint effects of predation and competition requires broadening of this concept to *interspecific density dependence* where increasing the density of one species affects the per capita growth rate of another species. Consistent with this usage, the traditional density dependence concept of ecology becomes *intraspecific density dependence*. Competition and predation both contribute to intraspecific and interspecific density dependence [17].

Basic understanding of the role of competition in ecological communities is usually represented by the Lotka–Volterra competition equations. Using the species labels to mean also their densities, these equations can be written in the form

$$\frac{1}{N_i}\frac{dN_i}{dt} = r_i\left(1 - \sum_{j=1}^{n}\alpha_{ij}N_j\right), i = 1, ..., n. \tag{13.1}$$

These equations define the per capita growth rate of each species in the focal guild, which reflect the average conditions that individuals of each species experience. The quantity r_i is the maximum per capita growth rate of species i, which is reduced by the terms representing density dependence inside the parenthesis. The coefficient α_{ij} measures density dependence of species j on species i. It measures how much the per capita growth rate of species i, as a proportion of its maximum value r_i, is decreased by increasing the density of species j by one unit. This coefficient measures interspecific density dependence if j is different from i, and intraspecific density dependence if $i = j$. The coefficient of intraspecific density dependence defines the so-called carrying capacity for a species: $K_i = 1/\alpha_{ii}$. Traditionally, the Lotka–Volterra competition equations have been parameterized with the competition coefficients defined as multiples of the carrying capacity, with the carrying capacity itself appearing explicitly in the equations, but this approach is now known to obscure the workings of the equations [15], and so is not done here.

Recent understanding shows that the Eq. 13.1 can also represent apparent competition, not just competition, and indeed they can represent the combined effects of competition and apparent competition. To do this, the equations are derived from a larger set of equations that take account of the direct interactions of the focal species with other species in the food web [17]. Thus, these equations have explicit formulae defining the effects of the arrows for a food web like Fig. 13.1. Equation 13.1 is then able to define the long-term growth of the species in the focal guild in terms of feedback loops from other species. The α_{ij} are given by formulae derived from this larger set of equations and depend specifically on how species forage for their resources, how they are preyed upon by their predators, and how these resources and predators change in density in response to density changes in the focal guild [17]. It is important to note, however, that Eq. 13.1 does not apply to short-term prediction, but to long-term outcomes such as long-term recovery of a population from low density, or eventual extinction [17]. Although the Lotka–Volterra competition equations are used generically in ecology, it should be appreciated that they are very specific, and can only be used to illustrate general principles, not to give precise outcomes in any given system. Fortunately, there are means of matching Lotka–Volterra models with models for particular systems that validate their use for general principles when treated appropriately [4, 15, 23].

The key issue to be addressed with the Eq. 13.1 is when species coexistence will occur, and when some species will be excluded from the community. This is easy and straightforward in the case of a guild of two species, and this case gives some key general principles [17]. The fundamental result in the two-species case is that species j can exclude i from a community if

$$\alpha_{ij} > \alpha_{jj}. \tag{13.2}$$

The reverse inequality means that species i is always able to invade the system when dominated by species j. The mutual invasibility criterion [24] then says that two species, labeled 1 and 2, will coexist stably whenever they are both able to recover from low density in the presence of the other species. In the two-species Lotka–Volterra equations, this criterion leads to the condition

$$\alpha_{11} > \alpha_{21} \text{ and } \alpha_{22} > \alpha_{12}. \tag{13.3}$$

Fundamentally, this means that for stable coexistence, each species must depress its own growth more strongly than it depresses the growth of the other species as it increases in population density. It is a very simple and general criterion that ensures stable coexistence [15]. If one of the inequalities in (Eq. 13.3) is reversed, then one species can exclude the other, and not vice versa. This means that one species always drives the other extinct. On the other hand, if both inequalities are reversed, then each species can exclude the other. This means that neither species can invade a system consisting of the other species. Whichever species establishes first remains the sole occupant in the guild in question in that locality.

The Lotka–Volterra competition Eq. 13.1 can be interpreted directly as meaning direct interference of individuals of all species with individuals of other species, harming them by reducing foraging time or in some cases by cannibalism or *intraguild predation*, which refers to predation by one species in a guild by another species in that guild [25]. In this case of direct interference, resource shortages or predators need not have a role [9–11]. Indeed, it is in this form that the

Lotka–Volterra equations are in principle most accurate, given that they do not explicitly represent the dynamics of resources or natural enemies outside the guild in question. In these cases, they are valid for short-term as well as long-term predictions. In these cases of direct interference, conditions (Eq. 13.2) and (Eq. 13.3) then apply, respectively, to exclusion and coexistence, showing in particular that coexistence requires each species to interfere more strongly intraspecifically than interspecifically.

When the Lotka–Volterra equations are used to represent trophic interactions such as those of the food web Fig. 13.1, the coefficients of density dependence α can be defined in terms of quantities that summarize the interactions in the food web: niche overlap, ρ, and species-level average fitness, κ [17]. The niche overlap, ρ, between any pair of species is a measure of the relative strength of the density-dependent feedback between versus within species through resources and through predators. This quantity takes the value 1 when there is complete overlap, and zero when then there is no overlap. No overlap occurs when the members of the pair of species under consideration do not share resources and do not share predators. Thus, they do not have arrows to or from any of the same resources or predators in Fig. 13.1. With complete overlap, they have arrows to and from all of the same resources and predators, and these predators and resources are of the same relative importance for each species in the guild. If the resources and predators vary in importance for different species, but nevertheless both species under consideration are affected by them, then ρ will be between 0 and 1. Figure 13.2 gives various scenarios for different strengths of niche overlap.

The average fitness measure, κ, for any given species measures its ability to meet its energy needs and avoid predation when all species in the guild are at low density and thus not providing any feedback through density changes. This measurement is also expressed in special units, namely, in units of average sensitivity of the per capita growth rate of the species to changes in resources and predators. These quantities ρ and κ now relate to the coefficients of density dependence according to the relationship

$$\frac{\alpha_{ij}}{\alpha_{jj}} = \frac{\kappa_j}{\kappa_i}\rho. \tag{13.4}$$

Thus, the ratio of interspecific to intraspecific density dependence for species j's impact on species i is equal to the ratio of the fitness of species j to species i, multiplied by the overlap measure. This relationship is correct regardless of how many species are present in the guild in question. However, when there are just two species, the condition that the ratio (Eq. 13.4) be greater than 1 is the condition (Eq. 13.2) for species j to exclude species i. The ratio being less than 1 means that exclusion does not occur. These conditions in terms of the fitness ratio, multiplied by the overlap measure, are also strongly intuitive and instructive in terms of how stable coexistence occurs.

Consider the case of complete niche overlap, $\rho = 1$. Then the formula (Eq. 13.4) implies that whichever species has the larger fitness will exclude the other species.

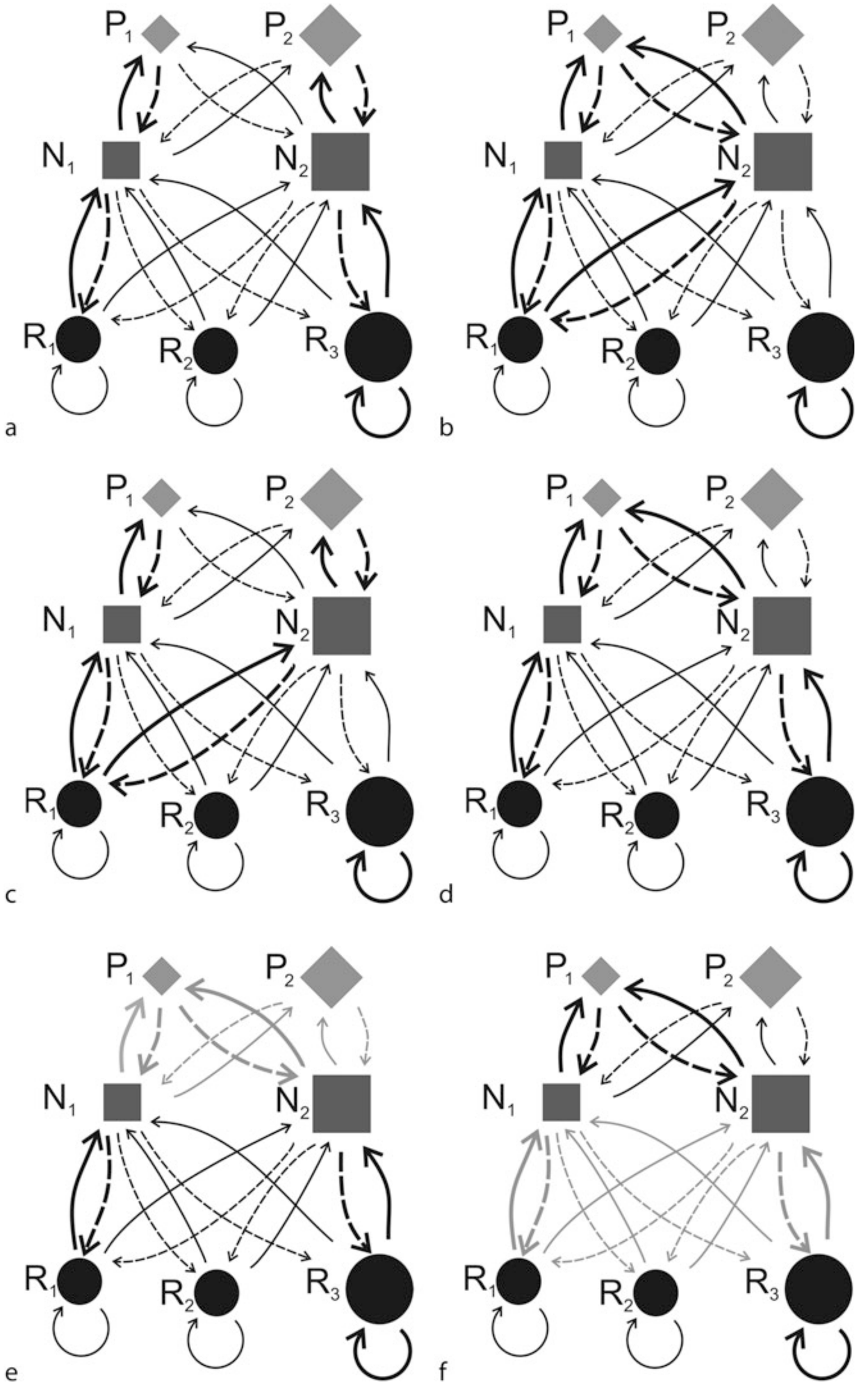


Fig. 13.2 Different scenarios for indirect linkages between the focal species, N_1 and N_2, through their resources and predators, leading to different degrees of niche overlap, ρ. Thick arrows

In other words, a simple fitness ratio greater than 1 is sufficient for one species to exclude the other. When $\rho < 1$, the fitness ratio is discounted by the overlap measure to determine if exclusion will occur. For example, if the overlap is ½, a species has to have fitness more than twice as high as the other to exclude it. The mathematics implies this outcome, but intuitively the idea is that the less species overlap in their niches, the more difficult it is for one species to exclude the other, but this can still occur if a species has a strong enough fitness advantage.

It is also worth noting that ρ is a symmetric measure of the ratio of interspecific to interspecific density dependence for any pair of species i and j, as formula (Eq. 13.4) implies that

$$\rho = \sqrt{\frac{\alpha_{ij}}{\alpha_{jj}} \cdot \frac{\alpha_{ji}}{\alpha_{ii}}}. \tag{13.5}$$

Thus, ρ can be thought of as niche overlap, and at the same time as the geometric mean of the ratios of interspecific versus intraspecific density feedback for any pair of species. Thus, it measures the potential for stabilization of coexistence without taking into account the fact the ratio might be different for one species compared with other and lead to dominance instead of coexistence. The formula (Eq. 13.4) shows how that bias changes as the fitness ratio changes, suggesting, as we shall see below, how changes in circumstances can change the opportunities for coexistence.

Niches and Species Coexistence Mechanisms

The results of the Lotka–Volterra equations can now be applied to see how a guild of coexisting species is constrained by trophic relationships. For two species to coexist, the conditions (Eq. 13.3) together with the formula (Eq. 13.4) imply that

$$\rho < \frac{\kappa_1}{\kappa_2} < \frac{1}{\rho}. \tag{13.6}$$

◄

Fig. 13.2 (continued) indicate relatively strong linkages compared with thin arrows. A gray arrow indicates that although the linkage might be important, it is not a strong source of density-dependent feedback for focal species. (**a**) Low niche overlap because the strong links to both resources and predators differ between N_1 and N_2. (**b**) High niche overlap because the strength of the links between N_1 and the various resources and predators has the same pattern as the strength of the links between N_2 and these resources and predators. Although these two species do discriminate between these resources and predators, they do so in an identical way. (**c**) High niche overlap through resources, but low niche overlap through predators, leading overall to an intermediate level of niche overlap. (**d**) Low niche overlap through resources, but high niche overlap through predators, leading overall to an intermediate level of niche overlap. (**e**) As in **d**, but predation is not strongly density dependent, and so overall niche overlap is low due to the low overlap through resources. (**f**) As in **d**, but resource consumption is not strongly density dependent, and so overall niche overlap is high due to the high overlap through predators

Fundamentally, this says that species 1 and 2 will coexist stably if their fitness ratio lies between ρ and $1/\rho$. Thus, the less the niche overlap, the greater the difference in fitness that is tolerated compatible with species coexistence. These conditions involve two kinds of differences between species that affect species coexistence in opposite ways. First are fitness differences, which are measured in terms of the deviation of the fitness ratios from 1. This ratio fundamentally measures the overall relative degrees of adaptedness of the species to their common environment. A species with a larger value of κ can be thought of as a better performer, and therefore inequality in this respect understandably favors exclusion.

A complication arises due to the fact that when different focal species are present, the resource species or predator species maintained in the food web may be different [17]. When we ask if species 1 can exclude species 2, those resources and predators present when species 1 is present alone are used to calculate the fitnesses and niche overlap for both species 1 and 2. When we ask if species 2 can exclude species 1, a different set of resources and predators might be present. This means that in some cases, the ρ and κ ratio for each end of the inequality in (Eq. 13.6) will be different, and will have to be calculated based on which focal species is present. While, this complication does not alter the fundamentals as to whether a given species can exclude another species from a community, it is important to keep in mind that ρ and the κ's may change with the circumstances [26, 27].

What do these critical quantities measure? Although the κ's and ρ both involve aspects of resource consumption and predation, they measure independent aspects. In particular, the quantity ρ is independent of how well the species are adapted to the environment. Instead it compares species in terms of which trophic links (links to resources and predators) are most important to them, and how important they are, as illustrated in Fig. 13.2. Importance is measured in terms of the ability of that link to generate density-dependent feedback [17]. This comparison shows how much the species interact with each other through their resources and natural enemies and so how much conflict there is between them: the larger ρ is, the larger the conflict. Naturally, two individual organisms have greater similarity and therefore greater conflict through trophic relationships within species than between species (not counting the effects of overall fitness differences), and ρ is a relative measure comparing between-species interactions through trophic links to within-species interactions. In other words, it compares the strength of interspecific feedback loops with intraspecific feedback loops, in essence adjusted to equal overall fitness. Thus, the ratio κ_1/κ_2 and niche overlap ρ represent two different kinds of comparison between species.

Exclusion Principles

Recognition of these two separate kinds of ecological comparison between species resolves a conundrum about the competitive exclusion principle [28], which is often stated as "no two species can occupy the same ecological niche." It has been a problematic principle because often the niche is not defined precisely enough to give the principle content. Indeed, there are many and various definitions of the niche [29], but perhaps the one most usefully behind the principle is the Eltonian niche which is about how an organism uses the environment [29]. However, when natural enemies are considered, the niche must include not just how a species uses the environment, but also how other organisms in the environment use that species. Thus, the niche needs to be defined as how an organism relates to the environment [15, 29]. Regardless of whether this idea can be measured in an absolute sense, the quantity ρ provides a relative sense in which two species relate to the environment. This definition specifically removes the overall level of adaptedness to the environment from the comparison between species, focusing on how they relate to it. This way of comparing niches differs markedly from what would be concluded from comparing niches according to the recent niche definition of Chase and Leibold [29]. Under their definition, the conditions in the environment defining zero per capita growth rates are the niche. That means species would only overlap completely if their growth rates were zero under exactly the same conditions. Species with $\rho = 1$ would overlap completely according to the approach here, but would not have zero growth under the same conditions unless their κ values were the same too.

The sense in which the competitive exclusion principle is correct is that no two species can coexist stably if their niches overlap completely, i.e., $\rho = 1$. The Lotka–Volterra approach defines this idea here, and makes it precise, but it is important to realize that it in fact emerges from a broad array of models, as will be discussed further below, without evidently any contrary models. More important, these ideas allow a quantitative approach to the competitive exclusion principle. A niche overlap value of $\rho =1$ is an unlikely occurrence in nature, but a value near 1 is not unreasonable. Here condition (Eq. 13.6) shows that when species have high niche overlap, their fitnesses are very closely constrained to be more nearly equal. Thus, there is a more quantitative exclusion principle that relates not just to competition, but to apparent competition too, and states that species with high niche overlap must have correspondingly high similarity in average fitness if they are to coexist. This principle then replaces a statement with limited application to a more significant one about the difficulty of coexistence for species with strongly overlapping niches. That degree of difficulty is measured by how similar in average fitness the species must be to allow coexistence with that degree of overlap.

Stabilizing and Equalizing Mechanisms

The expanded exclusion principle defined here is broadened with the realization that there are two general but not equivalent ways in which species coexistence can be favored. First is low overlap in niches, and second is similarity in average fitness. Mechanisms that lead to low niche overlap are called stabilizing mechanisms, and those that lead to similar average fitnesses are termed equalizing mechanisms. The issue that distinguishes these two ways of achieving coexistence is the role of feedback loops. Niche overlap, ρ, is fundamentally about feedback loops. Low niche overlap means feedback loops between species are weak relative to those within species. As emphasized above, this is essential for coexistence to be stabilized. Thus, mechanisms that bring about low niche overlap are called stabilizing mechanisms. Stabilizing mechanisms vary from the obvious to the subtle. Specialization of the members of a guild on different resources is the stabilizing mechanism termed resource partitioning [30], which quite directly leads to low contributions to ρ (Fig. 13.2a, d, and e). Likewise, if the natural enemies of the guild members are specialists, feedback loops through predation are separated (Fig. 13.2a and c). This is natural enemy partitioning [17]. As we shall see below, these direct and obvious stabilizing mechanisms are far from the only ones. Likely involved in these scenarios are trade-offs that provide advantages to specialization. For instance, consumers well equipped to exploit a particular resource, or predators well equipped to attack a particular species, may not be so well equipped to exploit a different resource or attack a different species because the very equipment that works well in a specific situation does not work so well in another [11, 31].

Equalizing mechanisms do not have to involve the feedback loops at all. In the Lotka–Volterra model considered here, the fitnesses are measured at low density, and so density feedback has no direct role [17]. One can ask what mechanisms might lead species to be similar in average fitness in a given environment. It is clear that the laws of physics come in at the ultimate level and constrain performance differences between species. But there are still numerous ways in which species might differ in efficiency at a given task. Many trade-offs might be seen as equalizing mechanisms in that doing well in one respect might mean doing less well in another respect [29]. For instance, defense against predation or harsh physical environmental conditions might lead to lower growth rates of individual organisms and perhaps slower rates of reproduction. Thus, one species might have higher survival rates but suffer in reproduction relative to another species, thus limiting the fitness differences between species that are possible.

There is as yet no general theory of equalizing mechanisms, but one is likely to emerge from general principles of community assembly and natural selection. Natural selection drives species to the limits of what is possible: for instance, not being defended against harsh conditions, while not growing fast, are certainly possible in an organism, but if it is also possible through a genetic change to grow faster if expenditures on defense are low, or to be defended if growth is slow, then that character is likely to evolve. The constraint on what is possible is approached, which enforces the trade-off [4]. This process happens within species, as it involves natural selection at the individual level. The process of relevance between species is community assembly. More efficient species arriving in a particular locality are likely to displace others there, and that process will continue until constraints on efficiency are approached, trade-offs apply, and fitness differences are minimized.

It should be recognized that many constraints in nature may have both stabilizing and equalizing aspects to them. Trade-offs associated with resource partitioning may lead to stabilization as well as equalization, provided similar profit is derived from use of the various different resources involved. There is also a body of thought leading to the neutral theory of community ecology which says that highly diverse natural communities, such as tropical forests, behave as if all average fitnesses are equal, and stabilizing mechanisms are absent [32]. In terms of the theory here, this would mean that the average fitnesses, κ, are the same for all species and $\rho = 1$ for every species. Coexistence cannot be stabilized in these circumstances: invader growth rates are zero, and in the Lotka–Volterra model, the total abundance of all species is stabilized, but relative abundances are not. They are instead predicted to drift as a consequence of the chance processes of individual birth and death. However, there is no general reason to expect this extreme situation to arise in nature. The patterns that agree with those in nature can also be predicted by models with stabilizing mechanisms present [33], and fitness differences in systems without stabilizing mechanisms quickly lead to the collapse of diversity [34].

Competition-Based and Predation-Based Coexistence Mechanisms

Competition is often thought of as a major factor that limits species diversity, without a clear distinction being made between interspecific and intraspecific competition. That thought quickly leads to the idea that lower population densities will lessen the magnitude of competition and therefore promote the maintenance of species diversity [16]. In particular, that line of reasoning leads to the idea that high mortality rates, or harsh and stressful conditions, might generally promote species diversity [35]. However, as emphasized above, it is the ratio of interspecific competition to intraspecific competition that is critical to species coexistence and not the absolute value. Thus, lowering densities, and indeed lowering the intensity of competition, need not have any effect on the ratio of interspecific to intraspecific competition, and therefore need have not have any effect on species coexistence. In fact, competition can be important for stable coexistence. If competition is the only form of density dependence, then it is essential for stable coexistence. The challenge for species coexistence is not competition per se but interspecific competition that is strong relative to intraspecific competition. A pattern of species interactions that intensifies intraspecific competition relative to interspecific competition is a competition-based species coexistence mechanism [17].

Recent appreciation that density dependence from predation can play a similar role to competition leads to the idea of predation-based coexistence mechanisms. These are mechanisms that concentrate intraspecific apparent competition relative to interspecific apparent competition [17]. What then are the joint effects of competition and predation on species coexistence? Both natural enemy attack and resource competition are likely to be important in most guilds of coexisting species simply because it is difficult to avoid these phenomena. In discussions of species coexistence, predation has sometimes been viewed as important primarily as a cause of high mortality, and therefore not distinguished from harsh or stressful conditions in its effects [16]. However, as remarked above, such conditions need have no effect on species coexistence because they need not be associated with density-dependent effects, although they would alter species average fitness ratios if they were species-specific in effect.

The feedback loops expected with trophic interactions mean that predation would likely act in a density-dependent way. However, it is also possible that its density dependence might be weak relative to other sources and function on different spatial and temporal scales. For instance, predators that range over a much larger area than the focal guild might not be very responsive to the changes in the density of the focal guild. Predators might also live much longer and so change in density much more slowly than their prey in the focal guild, and they might also depend more heavily on species outside the focal guild and so not be so responsive to the focal guild even though they inflict mortality on them. Thus, it is not unreasonable to entertain the hypothesis that in some situations predation acts in an approximately density-independent way.

Density-independent predation would reduce the fitnesses κ of the species in the focal guild. From the two-species coexistence condition, it is clear that it would only affect coexistence if the ratio κ_1/κ_2 were affected, or in the multispecies case, if the ratio of any species' fitness to the guild average fitness were affected [15]. If fitnesses are unequal in the absence of predation, the same reduction for each fitness has the effect of making the fitness ratio deviate further from the value 1, which is more likely to lead to exclusion than coexistence. However, a trade-off that led to a disproportionately large reduction in fitness for the species with larger fitness ("selective predation on the competitive dominant") [16, 36] would make the fitness ratios more equal provided predation was not too strong. The requirement that the predation be not too strong leads to an intermediate predation prediction: predation promotes diversity for intermediate intensities of predation [16, 36].

Empirical studies have noted strong effects of predators in some guilds [37]. Of particular note are those consisting of sessile intertidal organisms, where the predator or herbivore is necessary for the maintenance of diversity [36, 38]. These predators are often termed keystone predators due to their essential role [38, 39]. Discussions of keystone predators often emphasize selective predation without regard for a density-dependent role of predation [37]. However, density-independent selective predation is purely equalizing and would be incapable of stabilizing coexistence in the absence of a separate stabilizing mechanism, such as resource partitioning [16].

It is reasonable to expect that a keystone predator would respond at least to some extent in a density-dependent way to its prey. A density-dependent but nonselective predator, i.e., one that does not affect the fitness ratio, would in fact have the effect of undermining competition-based coexistence, and thereby undermine coexistence overall. This would occur because such a predator adds equally to intraspecific and interspecific density dependence, diluting the effect of resource partitioning and increasing the value of ρ, weakening coexistence. A density-dependent but selective predator need not have negative effects on coexistence provided it is selective on the competitive dominant [40], but it would still be limited in its effects on coexistence without the presence of other stabilizing mechanisms, because alone it would not add the sort of structure to feedback loops able to separate many species [16, 35]. Indeed, the patterning of feedback loops would be quite limited. Lotka–Volterra theory in this case predicts that at most two species can coexist stably when species in the focal guild do not partition resources, and there is only one predator present, regardless of how the predator distributes its attacks [41].

Fundamentally, development of the keystone species concept has viewed predation and competition as very different kinds of interactions, and has not focused adequately on the true requirement that density dependence needs to be concentrated more strongly within species than between species to promote stable coexistence [16]. Simply reducing the magnitude of competition uniformly does not achieve this, and equalizing fitnesses alone does not achieve this in the absence of a separate and strong stabilizing mechanism [16]. Certainly it is conceivable that a keystone predator might focus its attacks in such a way that fitness is approximately equalized across species, which would achieve approximate neutrality, but there is no mechanism that makes this likely. Trade-offs between susceptibility to predation and competitive ability would work in the right direction [40], but there is no reason for these trade-offs to be sufficiently precise to yield approximate neutrality, which nevertheless would not stabilize diversity, just slow its loss. Though keystone predation can indeed be shown to have strong effects in some communities [16], the stabilizing mechanisms needed to make it effective have neither been identified nor sought.

The theory discussed above identifies predator partitioning as an important way in which predation can promote diversity. Indeed, predators and other natural enemies can have very strong effects on their prey populations, and it is not uncommon for predators to be relatively specialized [11]. In fact, if the predators of a guild partition their resources (the prey populations in the guild in question), it follows reciprocally that the prey are partitioning predators. Thus, predator partitioning should have about the same prevalence in nature as resource partitioning. If both predator partitioning and resource partitioning are present for any given guild (Fig. 13.2a), the reasonable expectation is that the strength of coexistence would be stronger than if only one of these were partitioned (Fig. 13.2c and d). Indeed, the evidence from models points in this direction [17]. However, this outcome applies when predators that partition their prey are substituted for predators that do not partition their prey. A situation of great importance in nature is removal, addition, or restoration of predators, as occurs with human activities, although human activities might sometimes substitute one kind of predator for another [42–44]. In this case, adding predators that partition the focal guild to a comparable extent to the partitioning of resources in the focal guild would not greatly change the strength of coexistence. The reason is that there would be little change in the ratio of interspecific to intraspecific density dependence. However, adding predators that do not partition would increase the ratio of interspecific to intraspecific density dependence, undermining resource partitioning and therefore undermining coexistence, as noted above.

These effects of adding or removing predators, or changing their properties, can be summarized in terms of changing values of ρ as the scenario changes. Note that ρ is the niche overlap measure, but is also a symmetric measure of the ratio of interspecific to intraspecific density dependence by Eq. 13.5. Defining the case under consideration in parentheses gives the following chain of inequalities:

$$\begin{aligned}\rho(R_{\mathrm{p}}, P_{\mathrm{p}}) &\approx \rho(R_{\mathrm{p}}) \approx \rho(P_{\mathrm{p}}) \\ &< \rho(R_{\mathrm{p}}, P_{\mathrm{np}}) \approx \rho(R_{\mathrm{np}}, P_{\mathrm{p}}) < \rho(R_{\mathrm{np}}, P_{\mathrm{np}}) \\ &= \rho(R_{\mathrm{np}}) = \rho(P_{\mathrm{np}}) = 1.\end{aligned} \tag{13.7}$$

Here R or P indicates the presence of resource competition or apparent competition, and the subscripts "p" and "np" indicate partitioned and not partitioned interactions. As smaller values of ρ mean stronger coexistence, the situations are ranked from strongest coexistence to no coexistence, from left to right. The approximate equality of $\rho(R_{\mathrm{p}})$ and $\rho(P_{\mathrm{p}})$ in (Eq. 13.7) is not a conclusion, but the assumption that resources and predators are partitioned about equally, specifying the scenario considered here. The rest of the inequalities and approximations are conclusions. When P or R is not listed in parentheses, predators or resources may still be present, but are not important sources of density dependence. For instance, strongly density-dependent predation can prevent resource competition from occurring even though resources are still consumed and contribute essentially to fitness.

When competition and apparent competition are both present, the value of ρ is intermediate between the values that occur when only one of these is present. Thus, when they are both partitioned in inequalities (Eq. 13.7), the value of ρ does not change much, regardless of whether competition and predation are both present, or only one is present (e.g., Fig 13.2a vs. 13.2e). When they are both present, but only one is partitioned, the value of ρ is necessarily higher than when they are both partitioned because then a smaller fraction of all density-dependent interactions are partitioned, reducing the distinction between interspecific and intraspecific density dependence (Fig. 13.2c and d vs. 13.2a). The reason is that there is less partitioning overall among the array of density-dependent interactions experienced. No partitioning leads to a value of ρ equal to 1 (Fig. 13.2b), regardless of which interactions are present, and therefore no possibility for stable coexistence.

The situation not considered in the inequalities (Eq. 13.7) is when there is partitioning between predation and competition (Fig. 13.3). In this case, although there is no partitioning of resources or predators, there is joint partitioning of them in that some species have strong predator feedback loops, being particularly susceptible to predation, and some species have strong resource feedback loops and therefore are particularly susceptible to resource competition [40]. This is a predation-competition trade-off. But without partitioning within these interactions, at most two species can coexist. The predator is selective in this case, and it is simply the keystone species case once again. It implies

$$\rho(R_{\mathrm{np}}, P_{\mathrm{np}}) < \rho(R_{\mathrm{np}}) = \rho(P_{\mathrm{np}}) = 1 \tag{13.8}$$

as illustrated in Fig. 13.3. Although theoretically interesting in that it leads to coexistence in the absence of resource partitioning, it is not a serious solution

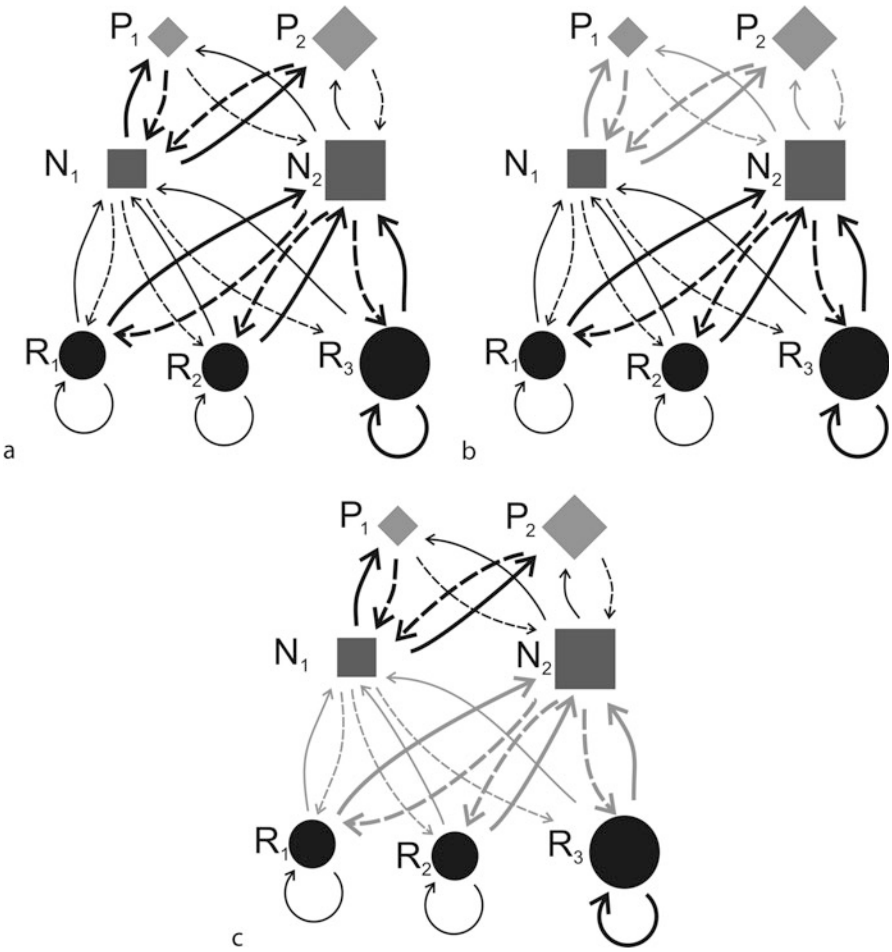


Fig. 13.3 General features as for Fig. 13.2. (**a**) Trade-off between competition and predation (apparent competition). Species 1 is more sensitive to predation, and species 2 is more sensitive to resources. Both resource consumption and predation are density dependent. Niche overlap is low. (**b**) As for (a), but only resource consumption is strongly density dependent. Niche overlap is high. (**c**) As for (a), but only predation is strongly density dependent. Niche overlap is high

to the maintenance of high diversity by means of predation because, by itself, it allows stable coexistence of at most two species as discussed in detail above.

Complexities of Foraging Behavior and Resource Acquisition

Predation and resource competition intimately involve the behavior of the organisms in seeking food and resources, or avoiding predation [45]. Interference competition is one behavior that is accounted for reasonably well in the

Lotka–Volterra competition equations, as discussed above, as direct contributions to the coefficients of density-dependence α. However, other aspects of the behavior of organisms are not accounted for well because all resource acquisition rates and foraging rates are assumed to be linear, i.e., the amount of any resource or prey consumed or captured by an individual organism per unit time is simply proportional to the abundance of that resource or prey [17]. One very simple and almost unavoidable deviation from the linearity of Lotka–Volterra equations occurs through the simple fact that an individual organism is generally limited in how much resource can be consumed per unit time, or how many prey can be captured and consumed. As a resource or prey becomes very abundant, the rate of consumption normally plateaus. Such plateaus weaken density-dependent feedback as densities increase, creating instabilities in population dynamics and promoting population fluctuations [1].

There is a vast literature on how these fluctuations might be stabilized and the extent to which they are indeed responsible for population fluctuations in nature [1]. Some scenarios involving population fluctuations provide further mechanisms of coexistence, because they enable population fluctuations to be partitioned by the various species, as discussed below (*environmental and population fluctuations*). However, when such nonlinearities are not strong enough to destabilize dynamics, and even in many cases when they do, the general principles for the Lotka–Volterra case continue to apply to the extent that they have been studied theoretically. For instance, a very general development of Levin [41], though lacking specific detail, is entirely consistent with the messages from the Lotka–Volterra development on partitioning resources and predators.

Resource-competition models where the resources do not regenerate in a Lotka–Volterra manner have also been studied [46], reproducing much of the detail, even quantitatively, for the Lotka–Volterra case. For plants, the key resources are light energy and simple inorganic compounds. Obtaining them in the right proportions is key to optimum plant growth, which makes the equations not just nonlinear, but nonadditive in the different resources. Despite these differences, the qualitative picture given here reappears in a different quantitative form [29, 31]. Nonlinear predation for equilibrium scenarios has been studied in limited situations [40, 47–49]. Quantitative variations on the results from the Lotka–Volterra case are found, but again in general the qualitative messages here remain intact. The reason is the fundamental nature of the requirement for coexistence that a species should inhibit its own growth more than it inhibits the growth of other species if it is to coexist with them. Thus, the requirement that intraspecific density dependence be stronger than interspecific density dependence is a robust requirement [15]. More complex nonlinear situations differ only in providing alternative ways of achieving this outcome. Partitioning of the environment, in one form or another, remains common among these alternative models. The major exception is for complex behaviors that lead to frequency dependence, as discussed next.

Some of the strongest effects occur when the per capita rates of foraging depend on the relative abundances of the species. The linear rates assumed in the Lotka–Volterra development mean that the fraction that any particular prey species represents in the diet of a predator is proportional to the abundance of that prey species. This makes intuitive sense as a prey species might be expected to be encountered at a rate proportional to its abundance. However, if a predator's efficiency at finding a prey species, or capturing it, increases with the rate at which it is encountered, then the actual rate of capture is super-proportional to the abundance of the prey species. The effect of this on the prey mortality rates is to make them increasing functions of intraspecific prey density, and possibly decreasing functions of the densities of other species. This phenomenon is known as frequency-dependent predation [50]. An important outcome is the intensification of intraspecific apparent competition relative to interspecific apparent competition. In models, frequency-dependent predation has been shown to be a powerful species coexistence mechanism. At a qualitative level, the outcome is similar to predator partitioning (or natural enemy partitioning) where each prey species has its own specialist natural enemy [50, 51]. Thus, frequency-dependent predation achieves with one natural enemy the same effect as several specialist natural enemies that do not have frequency-dependent behavior.

Frequency-dependent predation, which is also called "switching" in the ecological literature [52, 53] and "apostatic selection" in the evolutionary literature [54], has considerable empirical support to the extent that it has been studied [50]. Most studies reporting this phenomenon have been laboratory studies, with few prey species, and much less is known empirically from field studies. Moreover, the theoretical studies have only demonstrated multispecies coexistence from symmetrically acting forms of frequency-dependent predation: those that might be expected from increasing prey encounter and capture efficiency with increasing prey abundance [51]. The theory of optimal diet selection, however, leads to an asymmetrical form of frequency-dependent predation where the per capita predation rate on prey of low value to a predator depends on the absolute abundance of higher ranked prey, and not directly on the abundance of that prey itself [55]. Although coexistence can be promoted also by this form of behavior [56, 57], it is much less well investigated than symmetrically acting frequency dependence, and its performance in multispecies situations is not understood.

Environmental Variation and Population Fluctuations

Environmental fluctuations are often intuitively expected to undermine equilibrium perspectives such as those developed on the basis of Lotka–Volterra models [58]. In fact, however, rather than undermine them, they provide new contexts in which the critical phenomena considered here are realized. Fundamentally, they allow resource partitioning and predator partitioning to occur on large scales of time even though they do not occur on small scales [51, 59]. This is possible because different species in the same guild may have different responses to the physical environment, such as weather. Annual plant species, for instance, often have species-specific weather-dependent germination requirements. As weather patterns vary from year to year, and sometimes very markedly, this means that guilds of annual plants often show very marked variation from year to year in the relative abundances of the plants that successfully germinate, grow, and flower [59, 60]. Those that are not successful in given year can bide their time as dormant seeds in the soil, with good chances of success in some future year. In this way, intraspecific competition becomes concentrated relative to interspecific competition when a sequence of years of different environments are considered, as this behavior means that a given individual is likely to experience more intraspecific competition than interspecific competition from any given other species. The outcome is partitioning of resources over time [59]. Indeed, it is possible to derive Lotka–Volterra equations or very similar ones that represent the outcome of integration of short-term fluctuations over time, even based on random fluctuations on that short timescale [15, 23].

The details of these effects of temporal partitioning are important. It is critical for instance, that competition between individuals growing at different times does not occur [61]. For this to be the case, the resources cannot carry over in time. For instance, day and night foragers for a particular resource might still be competing strongly even though they capture the resource at different times. This would be the case if units of resource available at night are used by organisms in the day time if they are missed by night time foragers. The result is that temporal differences in foraging do not, in this case, concentrate intraspecific competition relative to interspecific competition.

What is possible in this context for competition is also possible with apparent competition, extending the essential symmetries for consideration of these processes before. However, differences do occur because of the potential that apparent competition works on a longer timescale [61]. For example, in annual plants with seed predators, when a good seed crop causes buildup of seed predators through high predator reproduction and survival, it is the seed crop the next year that experiences the higher predator density. The seed crop the next year will likely have different species composition due to species differences in response to the temporally varying physical environment. Thus, the species that cause the predator buildup are not necessarily the species experiencing higher predation. Intraspecific apparent competition is not necessarily concentrated relative to intraspecific competition in his case. When the environment varies randomly from 1 year to the next, predator buildup on good seed crops does not lead to effective temporal partitioning. However, behavioral changes in predation rates can be on short timescales with the potential that effective temporal partitioning can occur. Indeed, the theory implies that frequency-dependent behavior and other density-dependent behaviors can in some circumstances lead to temporal partitioning due to predation [50, 51]. In these cases, frequency-dependent behavior has two effects: an immediate and direct effect of frequency dependence on the ratio of intraspecific to interspecific density dependence, and a longer-term effect that involves an interaction between year to year temporal variation in the physical environment and temporal variation in foraging behavior as the prey species fluctuate in density in response to the environment.

These instances of temporal partitioning due to the responses of species to environmental variation are generally referred to as the storage effect, but another general mechanism termed relative nonlinearity of competition [15] works based on temporal fluctuations in the magnitude of competition and different nonlinear responses to competition by different species. Fundamentally, different species are more sensitive to competition over different ranges of competition. Under certain conditions, they can also generate the kinds of fluctuations that are disadvantageous to them when they are abundant. This mechanism continues the theme that coexistence stems from mechanisms that generate stronger intraspecific density dependence than interspecific density dependence when integrated over time. This mechanism works just as well for apparent competition [62], but it seems to be a relatively weak mechanism to the extent that at best two species can coexist this way in the absence of other mechanisms [15]. However, in the two-species case, models have suggested that it can have comparable strength to resource partitioning [63]. Moreover, there is some suggestion that when interacting with other mechanisms, it can be very powerful in multispecies settings, but there is as yet insufficient information on this possibility [64, 65].

Spatial Structure

Spatially varying environments have long been known to allow spatial partitioning of competition in a very similar way to temporal partitioning [66]. Thus, there are various spatial analogues to temporal partitioning by the storage effect for both competition [67–69] and apparent competition [22, 70]. Indeed, there is every indication that spatial variation powerfully fosters species coexistence due to the different responses that species have to environmental factors that vary in space. From time to time, the idea that species tend to be spatially aggregated intraspecifically is raised as a potential mechanism of species coexistence [71]. Such local aggregation naturally arises due to localized dispersal [72], can also be influenced by habitat selection. Although coexistence by aggregation has sometimes been claimed to be an entirely new mechanism [73], there are good reasons to believe that it functions primarily as a form of spatial resource partitioning [74], or as a form of competition–colonization trade-off [75], as discussed next.

Some spatial mechanisms do not have clear analogues with temporal mechanisms. Of particular note is the mechanism known as competition–colonization trade-offs [76]. In this case, species are assumed to be ranked very strictly in competitive ability such that any given locality will become dominated by just one species, the strongest competitor that has found that locality. In some cases, a locality is assumed to be a site only large enough to be occupied by a single plant individual. In other cases, it might be a local population of a plant or animal species [77]. Death of the individual, or extinction of the local population, vacates that space. The assumption is that there is a trade-off between competitive ability and colonizing ability, and this means that vacant space is likely to be taken by an inferior competitor. Ecological succession takes place as inferior competitors are replaced by later arriving superior competitors. Because local sites are vacated at different times, and recolonization and succession have stochastic timing, landscapes governed by these process will consist of a mosaic of sites in different successional stages. For this reason, this hypothesis is also known as the successional mosaic hypothesis [35]. It is one version of the intermediate disturbance hypothesis [78–80], and perhaps the version closest to the original intention of the idea [79].

In these discussions, disturbance is a natural process such as fire or extreme weather that destroys local populations patchily in space. In one version, predators are the agents that destroy local populations [37, 81], and thus maintain a patchy landscape in a mosaic of successional states. The competition–colonization trade-off hypothesis, however, can work without an agent of disturbance, but relying on chance mortality of individuals, dispersal and colonization [82]. This successional mosiac process can be modeled well by Lotka–Volterra competition equations with density measured at the landscape scale as the fraction of sites occupied by a species [76, 77, 83]. Its form is the same as an asymmetric interference competition model, but nevertheless governed by the coexistence conditions (Eq. 13.3) above, which mean that all species inhibit themselves more than they do other species. For superior competitors, this outcome occurs because inferior competitors are better at finding free species and so escape interspecific competition from superior competitors. This idea is also related to nonspatial models of exploitation of leftover resources, for example light not intercepted by a plant canopy, and so available to understorey species [84, 85]. Fundamentally, in the competition–colonization trade-off hypothesis and leftover resource models, superior competitors do not efficiently exploit all resources, leaving some to be exploited by those species that have lesser competitive ability but through a trade-off have achieved the ability to exploit the leftovers.

A final spatial mechanism involves natural enemies. Known as the Janzen-Connell hypothesis [86–88], it was originally proposed for tropical trees, but is closely related to the soil-feedback hypothesis for coexistence of species in grasslands [89]. The idea as applied to trees was that natural enemies specialized on a particular species would build up in abundance on or near a given tree. These natural enemies would then provide strong inhibition to the establishment of individuals of the same species there. Other species, however, would be able to establish. It is clear, however, that this is a form of natural enemy partitioning that does not require a spatial element, although it might well be enhanced by the spatial element. In the soil-feedback hypothesis, soil microorganism communities develop in the root zone of a particular individual plant. A preponderance of relatively species-specific harmful microorganisms leads to a net negative effect of establishment of the same species at that site, favoring others species to replace that individual or to thrive nearby [89, 90].

Applications

These ideas have applications in a number of other areas beyond the basic concept of how communities are structured, including invasion biology [91, 92], conservation biology [93, 94], and ecosystem functioning [95]. The focus here is on invasion biology. The biosphere is undergoing vast changes as a result of human activities. One activity is the introduction of new species to places where they were not previously found either deliberately or as an accidental by-product of commerce. Many transplanted species fail to perform well in a new environment, or if they do perform well enough to establish self-sustaining populations (to "naturalize"), they never become very abundant. However, a few species perform spectacularly well and become major pests or weeds, often displacing native species [96]. Although invasive species are not often responsible for regional extinction of native species, they can displace them locally and dramatically change the character of local communities. The local communities thus undergo a process of reassembly in response to the arrival of these invasive species. The ability of alien species to have these effects is often analyzed in terms of competition and predation, although other mechanisms, such as facilitation of one species by another can be expected to be important too [97]. The study of coexistence and exclusion mechanisms, as discussed here, fundamentally involves the question of whether a species can increase from very low density in the presence of other members of the guild, or to "invade." This invasibility analysis also applies to the question of how an alien invader successfully enters a local guild, and whether it displaces existing guild members.

The ability of an alien species to invade can be discussed in terms of the concept of niche opportunities [91]. Fundamentally, a niche opportunity means sufficient resources are available for the species in question to invade, given the risk from natural enemies that it will encounter. To some extent, lower risk from natural enemies can enable a species to invade at lower resource levels, as it would be able to do so at lower reproduction or survival based on those resources. A surfeit of resources is called a resource opportunity, while a low risk from natural enemies is an escape opportunity. The overriding question in invasion biology is why native species, in the eons of time, have not used up all opportunities to exploit a particular environment. There are a number of potential answers to this question.

First the local environment may have changed as a result of human activities or climate change, and the local community is therefore no longer well adapted to it, allowing the potential for a species from elsewhere to be better adapted than local species [91, 98]. This idea of change, however, should not be confused with a natural regime of disturbance or environmental variation that might be temporally partitioned by the native species. Like other persistent features of the environment, such environmental variation ought already be exploited by the native species, and so should not provide new opportunities. Unless, the natural regime of environmental variation has changed, there is no change that should be expected to facilitate invasion [98]. Pollution is an example of one common change that humans cause.

When this results in deposition of nitrogen in a nutrient-poor environment, it can have a large effect on local plant communities, facilitating the invasion of species adapted to exploit this resource [99]. The invasive species then potentially depress other resources that the native species depend on, harming them.

Second, the native community might have not existed for very long, or in the past was poorly connected to areas that might provide colonists or genetic diversity allowing adaptation to exploit all opportunities well. In a sense, such a local community is immature [91]. In these cases, niche opportunities exist, and species from elsewhere might have the adaptations allowing them to exploit the local environment more successfully than the native species. Islands that have been devastated by invasions might be in this category [96, 100], but so also might aquatic systems which suffer from the problem that dispersal from one place to another was historically much rarer than it is today. This idea has been applied on continental scale, suggesting that the Eurasian continent being larger contains species with stronger competitive ability than species in North America [101]. There is some evidence for this in plant species from Eurasian that use chemical interference successfully in competition with North American species [102]. The idea of biotic resistance says that localities with high species diversity ought to be more difficult to invade [103, 104], which makes sense to the extent that it means that more niche opportunities are foreclosed by the adaptations present in the more diverse community.

Third, invaders potentially arrive in the absence of their specialist natural enemies. Even if diseases, parasites, or predators are introduced with an invader at the same time, unless these natural enemies can exploit other species, they may well be lost early in the introduction because the invader in question was initially at too small a population size to sustain them. Such a species would have a strong natural enemy escape opportunity that would give it an advantage in a novel environment, especially if native species were attacked by natural enemies that do not affect the invader [97, 105]. Of course, native natural enemies do attack and inhibit invaders to varying extents, reducing their success [91, 97].

These various ideas can be analyzed within the Lotka–Volterra framework presented here [92, 106]. Fundamentally, if an invader has overall advantage such as novel competitive weapons for interfering with other species, is less susceptible to enemy attack, or is better overall in its adaptation to the environment, it will have an average fitness advantage over native species (κ ratios greater than 1), allowing it to invade, with the potential, if this advantage is large enough, to exclude native species, at least locally. Instead, it might not have an overall advantage, but simply an advantage under specific conditions, leading to low niche overlaps, ρ, with native species. This advantage might be that it can specialize better on a particular resource or particular environmental conditions, arising spatially or temporally, that native species are not fully exploiting. In many situations, it seems that elements of both are likely. It might for instance weakly partition resources or predators, but also have some fitness benefits, without average superiority over natives, but together this might be sufficient for invasion. On the other hand, the fact that even strongly successful invaders do not eliminate species in the same guild on a regional scale suggests that spatial partitioning, perhaps coupled with temporal partitioning, is sufficiently strong for natives to persist in a region even if they are eliminated or driven to low abundances locally [92, 107].

Invaders naturally interact with species outside their own guild also. As predators and diseases they have been more frequently the cause of extinctions of native species [44, 96]. However, they can also be the agents that enhance predators or diseases, damaging native species through apparent competition [43, 108]. Ways of controlling such invasives is therefore of substantial interest in conservation biology [43, 108].

Future Directions

The picture of community organization driven by competition and predation presented here has been developing for decades but it is relatively recently gelling on a new synthesis. It is now able to integrate competition, predation, temporal variation, and spatial structure in one common framework to gain understanding of not just individual mechanisms but their interactions as well [17, 51], as exemplified by the simple comparison presented above of how the strength of stabilization of coexistence changes with the circumstances. This picture now leads to comprehensive theoretical understanding that greatly clarifies a confusing picture of numerous potential mechanisms of a few years ago [109]. The most glaring lack now is in rigorous empirical study of mechanisms by which communities are structured, especially stable coexistence mechanisms [14].

Many empirical studies focus on features of mechanisms without truly testing whether they have a role in stabilizing diversity [13, 14]. Patterns of morphology of animals [11] and plants [110] that are related to how they gather resources or avoid predation can provide strong circumstantial evidence, but still leave open numerous possibilities as to the actual mechanisms [110]. One problem has been that the sort of data necessary to test mechanisms has not been clear because the mechanisms have not been truly understood, but that issue is now much less significant. One recent development is a theory of testing mechanisms [13] aimed to guide empirical studies. The key problem is to determine ways of testing whether density feedback loops do indeed concentrate intraspecific density dependence relative to interspecific density dependence in the manner proposed by a specific mechanism. In general, this is not an easy proposition, but it is also not infeasible for well-resourced project, given the right approach. Developing these approaches is the aim of the theory of testing mechanisms [13]. For example, to test for stabilization by temporal or spatial resource partitioning, the concept of covariance between environment and competition has been developed, which specifically measures how competition is linked to environmental factors [111]. Testing whether coexistence is stabilized by partitioning environmental conditions involves showing that covariance between environment and competition weakens as a species is reduced to low density in presence of unmanipulated competitors. More developments along these lines are needed to test other specific mechanisms.

New theoretical understanding has also revealed relatively basic deficiencies in empirical studies. For instance, frequency-dependent predation shows much promise as a strong coexistence mechanism, yet there are few rigorous field studies of this phenomenon [50]. It seems that the dominance of predation–competition trade-offs, including keystone predation ideas, has led to the neglect of frequency-dependent predation as a coexistence mechanism in empirical studies. However, it is also true that studies of keystone predation assumed that the mechanism was powerful when acting alone, and this impression seems likely to have inhibited the search for the stabilizing mechanisms that must be present also for keystone predation to be effective in promoting multispecies coexistence.

Naturally, the theory, though rich and not rigorously tested, still has some glaring deficiencies. One issue that complicates the Lotka–Volterra theory presented above is that as parameters are changed gradually, species in the resource and predator guilds that interact with the focal guild may not be supportable [27]. This possibility leads to abrupt changes in the coefficients of density dependence and related parameters [26]. It is also true that depressing a given member of the focal guild to low density may mean that only a subset of the resource species or predator species is present [17]. These facts, however, do not alter the invasion criteria presented here provided they are based on the resources and predators actually present in a given invasion scenario. They do, however, complicate interpretation of the criteria as parameters are changed. It is not known how important these issues are, and a comprehensive theory of their effects is needed.

The integrated understanding of the roles of predation and competition in species coexistence and exclusion presented here also raises the critical question of what factors control density dependence through these two processes. Of potentially major significance are other species in food webs beyond those directly linked to the focal guild. For instance, if the major natural enemies of the focal guild are themselves subject to strong density dependence from their natural enemies, they will be limited in their ability to change in abundance in response changes in the density of the focal guild. This means that density-dependent feedback to the focal guild through their natural enemies would be limited to behavioral responses of the natural enemies to focal guild densities. The idea of trophic cascades has long postulated how various density-dependent effects permeate from one part of a food web to another [112–114]. Integration of trophic cascade research with species coexistence research has the potential to make important advances with major implications for understanding the broader impacts that humans are having on the planet through widespread disruption of food webs, especially the destruction of many large carnivorous species [115–117].

Most of the discussion here has been purely ecological focused on interactions between a few species. However, natural populations and communities are shaped by evolution and community assembly processes, of which the considerations discussed above form just a small part. It is now known that evolutionary change can be fast to the extent that ecological dynamics and evolutionary dynamics cannot be completely separated [118]. The field of adaptive dynamics tries to integrate ecological and evolutionary process, but nevertheless still tends to treat these two processes as functioning on different timescales [119]. At the same time, much study of species coexistence focuses on the conditions that allow species coexistence, without considering the processes by which species assemble in communities. A proper treatment of community assembly would include challenging a local community with invaders from the pool species present regionally. There have been only limited theoretical studies of this sort [83, 120, 121]. Without an adequate treatment of these processes of adaptation and assembly, there is no true prediction of the structure of natural communities, and a seriously incomplete understanding of the role of competition and predation in community structure.

Finally, future theoretical research will likely take on the challenge of long-term climate change. The planet faces relative rapid climate change at the hand of human activities, but long-term climate change has always been a feature of the environment. However, theoretical models normally assume that climate fluctuations have stable long-term frequencies. There is a critical need to strip away this assumption, and develop theory that allows predictions even though the climate is not statistically stable. A useful theory would couple long-term climate change with the ability of populations to move on a spatially structured landscape as climate shifts change the viability of parts of their habitat. With such migration, there is the potential that climate fluctuations realized by species will have stable long-term frequencies as species track the shifting climate, but there are bound to be numerous new issues arising due to the fact that different species are likely to track the environment at different rates and in different ways [122].

A few years ago, imagining developments in the directions discussed here would have been daunting given the challenges that simpler theory gave. However, recent progress augers major extensions beyond the current limited contexts.

Acknowledgments I am grateful for comments on the manuscript by Jonathan Levine and for support from the National Science Foundation grant numbers DEB-0717222 and DEB-0816231.

Bibliography

Primary Literature

1. Murdoch WW, Briggs CJ, Nisbet RM (2003) Consumer-resource dynamics. Princeton University Press, Princeton, NJ
2. MacArthur R (1970) Species packing and competitive equilibrium for many species. Theor Popul Biol 1:1–11
3. MacArthur RH (1972) Geographical ecology: patterns in the distribution of species. Harper & Row, New York
4. Tilman D (1982) Resource competition and community structure. Princeton University Press, Princeton, NJ

5. Grubb PJ (1977) The maintenance of species-richness in plant communities: the importance of the regeneration niche. Biol Rev 52:107–145
6. Harper JL (1977) Population biology of plants. Academic, London
7. Connolly SR, Roughgarden J (1999) Theory of marine communities: competition, predation, and recruitment-dependent interaction strength. Ecol Monogr 69:277–296
8. Frank DA, McNaughton SJ, Tracy BF (1998) The ecology of the Earth's grazing ecosystems. Bioscience 48:513–521
9. Birch LC (1957) The meanings of competition. Am Nat 91:5–18
10. Schoener TW (1983) Field experiments on interspecific competition. Am Nat 122:240–285
11. Dayan T, Simberloff D (2005) Ecological and community-wide character displacement: the next generation. Ecol Lett 8:875–894
12. Adler PB, HilleRisLambers J, Levine JM (2007) A niche for neutrality. Ecol Lett 10:95–104
13. Chesson P (2008) Quantifying and testing species coexistence mechanisms. In: Valladares F, Camacho A, Elosegui A, Gracia C, Estrada M, Senar JC, Gili JM (eds) Unity in diversity: reflections on ecology after the legacy of Ramon Margalef. Fundacion BBVA, Bilbao, pp 119–164
14. Siepielski AM, McPeek MA (2010) On the evidence for species coexistence: a critique of the coexistence program. Ecology 91:3153–3164
15. Chesson P (2000) Mechanisms of maintenance of species diversity. Annu Rev Ecol Syst 31:343–366
16. Chase JM, Abrams PA, Grover JP, Diehl S, Chesson P, Holt RD, Richards SA, Nisbet RM, Case TJ (2002) The interaction between predation and competition: a review and synthesis. Ecol Lett 5:302–315
17. Chesson P, Kuang JJ (2008) The interaction between predation and competition. Nature 456:235–238
18. Holt RD (1977) Predation, apparent competition, and the structure of prey communities. Theor Popul Biol 12:197–229
19. Nicholson AJ (1937) The role of competition in determining animal populations. JSIR (Australia) 10:101–106
20. Hutchinson GE (1959) Homage to Santa Rosalia or why are there so many kinds of animals. Am Nat 93:145–159
21. Roughgarden J (1989) Community structure and assembly. In: Roughgarden J, May RM, Levin SA (eds) Perspectives in ecological theory. Princeton University Press, Princeton, NJ, pp 203–226
22. Holt RD (1984) Spatial heterogeneity, indirect interactions, and the coexistence of prey species. Am Nat 124:377–406
23. Chesson P, Huntly N (1993) Temporal hierarchies of variation and the maintenance of diversity. Plant Species Biol 8:195–206
24. Kang Y, Chesson P (2010) Relative nonlinearity and permanence. Theor Popul Biol 78:26–35
25. Holt RD, Polis GA (1997) A theoretical framework for intraguild predation. Am Nat 149:745–764
26. Abrams PA (1998) High competition with low similarity and low competition with high similarity: exploitative and apparent competition in consumer-resource systems. Am Nat 152:114–128
27. Abrams PA, Rueffler C (2009) Coexistence and limiting similarity of consumer species competing for a linear array of resources. Ecology 90:812–822
28. Hardin G (1960) The competitive exclusion principle. Science 131:1292–1297
29. Chase JM, Leibold MA (2003) Ecological Niches: linking classical and contemporary approaches. The University of Chicago Press, Chicago
30. Schoener TW (1974) Resource partitioning in ecological communities. Science 185:27–39
31. Tilman D (1988) Plant strategies and the dynamics and structure of plant communities. Princeton University Press, Princeton, NJ

32. Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ
33. Purves DW, Pacala SW (2005) Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In: Burslem D, Pinard M, Hartley S (eds) Biotic interactions in the tropics. Cambridge University Press, Cambridge, UK, pp 103–138
34. Zhang DY, Lin K (1997) The effects of competitive asymmetry on the rate of competitive displacement: how robust is Hubbell's community drift model? J Theor Biol 188:361–367
35. Chesson P, Huntly N (1997) The roles of harsh and fluctuating conditions in the dynamics of ecological communities. Am Nat 150:519–553
36. Lubchenco J (1978) Plant species diversity in a marine intertidal community: importance of herbivore food preference and algal competitive abilities. Am Nat 112:23–39
37. Gurevitch J, Morrison JA, Hedges LV (2000) The interaction between competition and predation: a meta-analysis of field experiments. Am Nat 155:435–453
38. Paine RT (1969) A note on trophic complexity and community stability. Am Nat 103:91–93
39. Power ME, Tilman D, Estes JA, Menge BA, Bond WJ, Mills LS, Daily G, Castilla JC, Lubchenco J, Paine RT (1996) Challenges in the quest for keystones. Bioscience 46:609–620
40. Grover JP, Holt RD (1998) Disentangling resource and apparent competition: realistic models for plant-herbivore communities. J Theor Biol 191:353–376
41. Levin SA (1970) Community equilibria and stability, and an extension of the competitive exclusion principle. Am Nat 104:413–423
42. Crooks KR, Soule ME (1999) Mesopredator release and avifaunal extinctions in a fragmented system. Nature 400:563–566
43. Johnson CN, Isaac JL, Fisher DO (2007) Rarity of a top predator triggers continent-wide collapse of mammal prey: dingoes and marsupials in Australia. Proc R Soc B-Biol Sci 274:341–346
44. Knowlton JL, Donlan CJ, Roemer GW, Samaniego-Herrera A, Kertt BS, Wood B, Aguirre-Munoz A, Faulknier KR, Tershy BR (2007) Eradication of non-native mammals and the status of insular mammals on the California Channel Islands, USA, and Pacific Baja California Peninsula Islands, Mexico. Southwest Nat 52:528–540
45. Stephens DW, Brown JS, Ydenberg RC (eds) (2007) Foraging: behavior and ecology. University of Chicago Press, Chicago
46. Haygood R (2002) Coexistence in MacArthur-style consumer–resource models. Theor Popul Biol 61:215–223
47. Holt RD (1985) Density-independent morality, nonlinear competitive interactions and species coexistence. J Theor Biol 116:479–493
48. Kotler BP, Holt RD (1989) Predation and competition: the interaction of two types of species interactions. Oikos 54:256–260
49. Holt RD, Grover J, Tilman D (1994) Simple rules for interspecific dominance in systems with exploitative and apparent competition. Am Nat 144:741–771
50. Kuang JJ, Chesson P (2010) Interacting coexistence mechanisms in annual plant communities: frequency-dependent predation and the storage effect. Theor Popul Biol 77:56–70
51. Chesson P, Kuang JJ (2010) The storage effect due to frequency-dependent predation in multispecies plant communities. Theor Popul Biol 78:148–164
52. Murdoch WW, Bence J (1987) General predators and unstable prey populations. In: Kerfoot WC, Sih A (eds) Predation: direct and indirect impacts on aquatic communities. University Press of New England, Hanover and London, pp 17–30
53. Abrams PA, Matsuda H (2004) Consequences of behavioral dynamics for the population dynamics of predator-prey systems with switching. Popul Ecol 46:13–25
54. Bond AB (2007) The evolution of color polymorphism: crypticity, searching images, and apostatic selection. Annu Rev Ecol Evol Syst 38:489–514
55. Stephens DW, Krebs JR (1986) Foraging theory. The Princeton University Press, Princeton, NJ

56. Hamback P (1998) Seasonality, optimal foraging and prey coexistence. Am Nat 152:881–895
57. Krivan V (2003) Competitive co-existence by adaptive predators. Evol Ecol Res 5:1163–1182
58. May RM (1974) Stability and complexity in model ecosystems, 2nd edn. Princeton University Press, Princeton, NJ
59. Chesson P, Gebauer RLE, Schwinning S, Huntly N, Wiegand K, Ernest MSK, Sher A, Novoplansky A, Weltzin JF (2004) Resource pulses, species interactions and diversity maintenance in arid and semi-arid environments. Oecologia 141:236–257
60. Angert AL, Huxman TE, Chesson P, Venable DL (2009) Functional tradeoffs determine species coexistence via the storage effect. Proc Natl Acad Sci USA 106:11641–11645
61. Kuang JJ, Chesson P (2009) Coexistence of annual plants: generalist seed predation weakens the storage effect. Ecology 90:170–182
62. Kuang JJ, Chesson P (2008) Predation-competition interactions for seasonally recruiting species. Am Nat 171:E119–E133
63. Abrams P, Holt R (2002) The impact of consumer–resource cycles on the coexistence of competing consumers. Theor Popul Biol 62:281–295
64. Huisman J, Weissing FJ (1999) Biodiversity of plankton by species oscillations and chaos. Nature 402:407–410
65. Huisman J, Weissing FJ (2001) Fundamental unpredictability in multispecies competition. Am Nat 157:488–494
66. Amarasekare P, Nisbet RM (2001) Spatial heterogeneity, source-sink dynamics, and the local coexistence of competing species. Am Nat 158:572–584
67. Muko S, Iwasa Y (2003) Incomplete mixing promotes species coexistence in a lottery model with permanent spatial heterogeneity. Theor Popul Biol 64:359–368
68. Snyder RE, Chesson P (2004) How the spatial scales of dispersal, competition, and environmental heterogeneity interact to affect coexistence. Am Nat 164:633–650
69. Miller AD, Chesson P (2009) Coexistence in disturbance prone communities: how a resistance-resilience trade off generates coexistence via the storage effect. Am Nat 173:E30–E43
70. Amarasekare P (2008) Spatial dynamics of foodwebs. Annu Rev Ecol Evol Syst 39:479–500
71. Ives AR (1988) Covariance, coexistence and the population dynamics of two competitors using a patchy resource. J Theor Biol 133:345–361
72. Murrell DJ, Purves DW, Law R (2001) Uniting pattern and process in plant ecology. Trends Ecol Evol 16:529–530
73. Hartley S, Shorrocks B (2002) A general framework for the aggregation model of coexistence. J Anim Ecol 71:651–662
74. Chesson P, Neuhauser C (2002) Intraspecific aggregation and species coexistence. Trends Ecol Evol 17:210–211
75. Bolker B, Pacala S (1999) Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. Am Nat 153:575–602
76. Tilman D (1994) Competition and biodiversity in spatially structured habitats. Ecology 75:2–16
77. Hastings A (1980) Disturbance, coexistence, history, and competition for space. Theor Popul Biol 18:363–373
78. Connell JH (1978) Diversity in tropical rain forests and coral reefs. Science 199:1302–1310
79. Connell JH (1979) Tropical rainforests and coral reefs as open non-equilibrium systems. In: Anderson RM, Turner BD, Taylor LR (eds) Population dynamics. Blackwell Scientific Publications, Oxford, pp 141–163
80. Roxburgh SH, Shea K, Wilson JB (2004) The intermediate disturbance hypothesis: patch dynamics and mechanisms of species coexistence. Ecology 85:359–371
81. Caswell H (1978) Predator-mediated coexistence: a nonequilibrium model. Am Nat 112:127–154
82. Bolker BM, Pacala SW, Neuhauser C (2003) Spatial dynamics in model plant communities: what do we really know? Am Nat 162:135–148

83. Dushoff J, Worden L, Keymer J, Levin S (2002) Metapopulations, community assenbly, and scale invariance in aspect space. Theor Popul Biol 62:329–338
84. Kohyama T (1993) Size-structured tree populations in gap-dynamic forest–the forest architecture hypothesis for the stable coexistence of species. J Ecol 81:131–143
85. Vance RR, Nevai AL (2007) Plant population growth and competition in a light gradient: a mathematical model of canopy partitioning. J Theor Biol 245:210–219
86. Connell JH (1970) On the role of natural enemies in preventing competitive exclusion in some marine animals and rainforest trees. In: den Boer PJ, Gradwell G (eds) Dynamics of populations. Centre for Agricultural Publishing and Documentation, Wageningen, pp 298–312
87. Janzen DH (1970) Herbivores and the number of tree species in tropical forests. Am Nat 104:501–528
88. Adler FR, Muller-Landau HC (2005) When do localized natural enemies increase species richness? Ecol Lett 8:438–447
89. Bever JD, Dickie IA, Facelli E, Facelli JM, Klironomos J, Moora M, Rillig MC, Stock WD, Tibbett M, Zobel M (2010) Rooting theories of plant community ecology in microbial interactions. Trends Ecol Evol 25:468–478
90. Klironomos JN (2002) Feedback with soil biota contributes to plant rarity and invasiveness in communities. Nature 417:67–70
91. Shea K, Chesson P (2002) Community ecology theory as a framework for biological invasions. Trends Ecol Evol 17:170–176
92. Melbourne BA, Cornell HV, Davies KF, Dugaw CJ, Elmendorf S, Freestone AL, Hall RJ, Harrison S, Hastings A, Holland M, Holyoak M, Lambrinos J, Moore K, Yokomizo H (2007) Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? Ecol Lett 10:77–94
93. Drury WH (1998) Chance and change: ecology for conservationists. University of California Press, Berkeley, Calif
94. Sergio F, Caro T, Brown D, Clucas B, Hunter J, Ketchum J, McHugh K, Hiraldo F (2008) Top predators as conservation tools: ecological rationale, assumptions, and efficacy. Annu Rev Ecol Evol Syst 39:1–19
95. Kinzig A, Pacala S, Tilman D (eds) (2001) The functional consequences of biodiversity. Princeton University Press, Princeton, NJ
96. Sax DF, Gaines SD (2008) Species invasions and extinction: the future of native biodiversity on islands. Proc Natl Acad Sci USA 105:11490–11497
97. Reinhart KO, Callaway RM (2006) Soil biota and invasive plants. New Phytol 170:445–457
98. Sher AA, Hyatt LA (1999) The disturbed resource-flux invasion matrix: a new framework for patterns of plant invasion. Biological Invasions 1:107–114
99. Brooks ML (2003) Effects of increased soil nitrogen on the dominance of alien annual plants in the Mojave Desert. J Appl Ecol 40:344–353
100. Simberloff D (1995) Why do introduced species appear to devastate islands more than mainland areas? Pac Sci 49:87–97
101. Huston MA (1994) Biological diversity, 1st edn. Cambridge University Press, Cambridge
102. Callaway RM, Ridenour WM (2004) Novel weapons: invasive success and the evolution of increased competitive ability. Front Ecol Environ 2:436–443
103. Elton C (1958) The ecology of invasions by animals and plants. Methuen and Co, London
104. Levine JM (2000) Species diversity and biological invasions: relating local process to community pattern. Science 288:852–854
105. Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. Trends Ecol Evol 17:164–169
106. MacDougall AS, Gilbert B, Levine JM (2009) Plant invasions and the niche. J Ecol 97:609–615
107. Davies KF, Chesson P, Harrison S, Inouye B, Melbourne BA, Rice KJ (2005) Spatial heterogeneity explains the scale dependence of the native-exotic diversity relationship. Ecology 86:1602–1610

108. Courchamp F, Woodroffe R, Roemer G (2003) Removing protected populations to save endangered species. Science 302:1532
109. Wilson JB (2011) The twelve theories of co-existence in plant communities: the doubtful, the important and the unexplored. J Veg Sci 22:184–195
110. Ackerly DD (2003) Community assembly, niche conservatism, and adaptive evolution in changing environments. Int J Plant Sci 164:S165–S184
111. Sears ALW, Chesson P (2007) New methods for quantifying the spatial storage effect: an illustration with desert annuals. Ecology 88:2240–2247
112. Holt RD (2000) Trophic cascades in terrestrial ecosystems. Tree 15:444–445
113. Chase JM (2003) Strong and weak trophic cascades along a productivity gradient. Oikos 101:187–195
114. Terborgh J, Estes JA (2010) Trophic cascades: predators, prey, and the changing dynamics of nature. Island Press, Washington DC
115. Terborgh J, Lopez L, Nunez P, Rao M, Shahabuddin G, Orihuela G, Riveros M, Ascanio R, Adler GH, Lambert TD, Balbas L (2001) Ecological meltdown in predator-free forest fragments. Science 294:1923–1926
116. Myers RA, Baum JK, Shepherd TD, Powers SP, Peterson CH (2007) Cascading effects of the loss of apex predatory sharks from a coastal ocean. Science 315:1846–1850
117. Letnic M, Koch F (2010) Are dingoes a trophic regulator in arid Australia? A comparison of mammal communities on either side of the dingo fence. Austral Ecol 35:167–175
118. Meyer JR, Ellner SP, Hairston NG, Jones LE, Yoshida T (2006) Prey evolution on the time scale of predator-prey dynamics revealed by allele-specific quantitative PCR. Proc Natl Acad Sci USA 103:10690–10695
119. Geritz SAH, Kisdi E, Meszena G, Metz JAJ (1998) Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12:35–57
120. Morton RD, Law R (1997) Regional species pools and the assembly of local ecological communities. J Theor Biol 187:321–331
121. O'Dwyer JP, Lake JK, Ostling A, Savage VM, Green JL (2009) An integrative framework for stochastic, size-structured community assembly. Proc Natl Acad Sci USA 106:6170–6175
122. Davis MB (1986) Climatic instability, time-lags and community disequilibrium. In: Diamond J, Case T (eds) Community ecology. Harper and Row, Cambridge, pp 269–284

Books and Reviews

Eisenberg C (2010) The wolf's tooth: keystone predators, trophic cascades, and biodiversity. Island Press, Washington, DC

Grant PR, Grant BR (2008) How and why species multiply: the radiation of Darwin's finches. Princeton University Press, Princeton

Chapter 14
Species Diversity Within and Among Ecosystems

Jonathan E.M. Baillie and Karolyn Upham

Glossary

Biodiversity	The variability among living organisms from all sources and the ecological complexes of which they are part biodiversity includes diversity within species between species and of ecosystems.
Ecosystem	A dynamic complex of plant, animal, and microorganism communities and their nonliving environment interacting as a functional unit.
Ecosystem service	The benefits people obtain from ecosystems these include provisioning services such as food and water regulating services such as flood and disease control cultural services such as recreation or spiritual benefits and supporting services such as nutrient cycling.
Population	A geographic entity within a species that is distinguished either ecologically or genetically.
Species (biological species concept)	Interbreeding natural groups whose members are unable to successfully reproduce with members of other such groups.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

J.E. Baillie (✉) • K. Upham
Zoological Society of London, Regent's Park, NW1 4RY London, UK
e-mail: jonathan.baillie@zsl.org; Karolyn.Upham@zsl.org

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_14,

Definition of Species Diversity

Species diversity is a function of species richness, the number of species in a given locality and species evenness, the degree to which the relative abundances of species are similar [1, 2]. While this notion may be easy to conceptualize, it has proven difficult, and at time contentious, to quantify [1, 2]. Commonly used methods include constructing mathematical indices known as diversity indexes (the Shannon, Simpson, and Margalef indexes being the most widely used) or comparing observed patterns of species abundance to theoretical models [3]. There is no single best metric and often commonly used ones are chosen because they are familiar and not necessarily because they are the most appropriate [1]. The scale of assessment may range from within a single site or habitat (known as α diversity) to the difference between two or more sites (β diversity) [4], which can then be combined to give γ diversity – the diversity of the landscape. Early research tended to be concentrated on largely α diversity, but there has been a marked increase in studies investigating β and γ diversity during the last decade [1].

The species richness aspect of diversity is often considered the iconic measure of biodiversity since it is well defined and aligns with our intuitive sense of the concept [1, 5]. As the species level is the level at which living organisms are most widely known, species have become the major taxonomic rank for describing biodiversity. Along with populations, it is also the level that most scientists, managers, and policy makers use when referring to biological diversity, and it provides a useful means for both monitoring and priority setting [6]. As the species level is one of three fundamental levels of biodiversity (the others are the genetic level and the ecosystem/community level), species diversity is thus an important component of conservation. This entry will refer to species diversity in the less restrictive sense, that is, the variety of species that inhabit our planet.

Introduction

Ever since naturalists began classifying living organisms, there has been no consensus on how best to define a species [7]; Mayden [8] listed 24 different named species concepts and in a partial listing, Mallet [9] indentified 15. This longstanding failure of scientists to agree on how to identify species is commonly referred to as the species problem, with taxonomists being described as either “lumpers” or “splitters” with the former approach resulting in lower numbers of species and the latter higher. However, despite this proliferation of definitions, most taxonomists agree that species represent a distinct genetic lineage, interact with the environment in similar ways, and are reproductively compatible [1, 5, 7]. More importantly, whether based on gene flow, ecological separation, or morphological distinctiveness, most concepts tend to give similar results due to the independent evolutionary history that has resulted in unique morphological, ecological, and

Table 14.1 Estimates of number of eukaryote species globally, following [10]

Estimate	References	Method
30 million	Erwin [17]	Extrapolation from samples
3–5 million	Raven [18]	Ratios known: unknown species
10–80 million	Stork [19]	Extrapolation from samples
4.9–6.6 million	Stork and Gaston [20]	Ratios known: unknown species
1.84–2.57 million	Hodkinson and Casson [21]	Ratios known: unknown species
5 million	Hodkinson [22]	Ratios known: unknown species
4–6 million	Novotny et al. [23]	Extrapolation from samples
7.4–10 million	Mora et al. [16]	Extrapolation from samples

reproductive characteristics. In addition, most populations within a species share common ancestors with other populations in the very recent past, which in turn causes them to not be significantly differentiated [10].

Since the middle of the twentieth century, the most commonly accepted definition has been the biological species concept [10], which defines a species as a group of interbreeding natural populations whose members are unable to successfully reproduce with members of other such groups [11]. As species based on this definition have natural and objective boundaries due to gene flow [12], they provide natural units for biodiversity assessment [5]. While the approach of classifying species will influence both conservation priorities and the scale of intervention [5], the important thing for conservationists is that the commonly understood units can be defined and interventions implemented and assessed. There will always be issues with using species as a basis for biodiversity assessment, but many of these can be overcome [5, 10, 13].

The next question that often arises is how many species exist on Earth. While there are no reliable estimates for prokaryotes [10, 14, 15], many estimates exist for eukaryotic species. Recent studies [16] predict ~8.7 million species globally, of which ~2.2 million are marine species. Earlier estimates have varied greatly, with most falling between five and ten million (see Table 14.1). However, these estimates could be much higher if poorly known groups such as deep-sea organisms, algae, or fungi have more species than currently believed [10]. Incomplete sampling, lack of robust extrapolation approaches, controversy over the underlying assumptions, and subjectivity are all cited reasons for the uncertainty [16]. However, with only around 1.7 million species currently described [16] and only 15,000 newly described species each year [25], the key point is that a major knowledge gap still exists in this area; there is a likelihood of losing species to extinction before ever describing them.

Patterns of Species Diversity

Despite the uncertainty around how to define a species and the total number of species on the planet, the diversity of species is evident in the world around us. Species are unevenly distributed across the globe [10, 26] in terms of both geography and taxonomic groupings, with a disproportionately large number of species in the Class Insecta. It is very difficult to estimate this variability, as data are limited for marine and fresh water ecosystems and for non-vertebrate species, but there are around 10 plant species and at least 100 invertebrate species for each vertebrate species. Despite their low numbers compared to invertebrates, considerable datasets have been compiled for vertebrates, allowing numerous patterns of variations to be explored [10], although the extent to which these patterns can be generalized remains unknown [27].

One of the most obvious patterns is that for virtually all taxonomic groups, species richness tends to decrease from the tropics to the poles [28]; the tropics hold much higher species richness than do temperate, boreal, or polar regions. This is more so than would be expected on the basis of area alone [10], although it is complicated by physiographic and climatic factors such as mountain ranges or rainfall patterns. There is also a broadly similar spatial distribution of diversity between taxa, where differences seem to be driven by particular biological traits [10, 29]. In general, these differences will increase with increasing evolutionary distance between taxa [30]. Patterns of marine and freshwater species richness are less understood, but studies [31, 32] have demonstrated a latitudinal gradient in the shallow water benthos, with decreasing richness toward the poles.

There are also patterns of endemism and evolutionary distinctiveness that are important aspects of species diversity, as species with long and independent evolutionary histories and few surviving relatives contain irreplaceable genetic diversity. Most species have small range sizes [33], which tend to co-occur in centers of endemism [10, 31]. Centers of endemism are also concentrated in the tropics, with overlap, at certain scales, across birds, mammals, and amphibians [34] and a similar pattern is expected for plants [35]. However, this pattern does not appear to represent the situation for invertebrates or microorganisms [10]. There is enormous variation between species in terms of the evolutionary age [36]. Evolutionary distinctiveness among species can be explored using taxonomic relationships and as with species richness, the available data indicates that tropical rainforests are regions with the greatest number of taxa with the longest independent evolutionary history [37].

Vertebrate Species Status, Trends, and Threats

Conservation assessments have been primarily carried out for vertebrates as the conservation community tends to use these species as flagships or indicators of ecosystem health and global biodiversity. Therefore, this section will focus on vertebrate species, as sufficient data exists on the status and trends of the world's mammals, birds, amphibians, and cartilaginous fishes, and weighted approaches can be used to gain an understanding for reptiles and bony fishes [38]. Other taxonomic groups will not be discussed as there is not currently sufficient data for a global assessment. However, vertebrates represent only 5% of animal species and a fraction

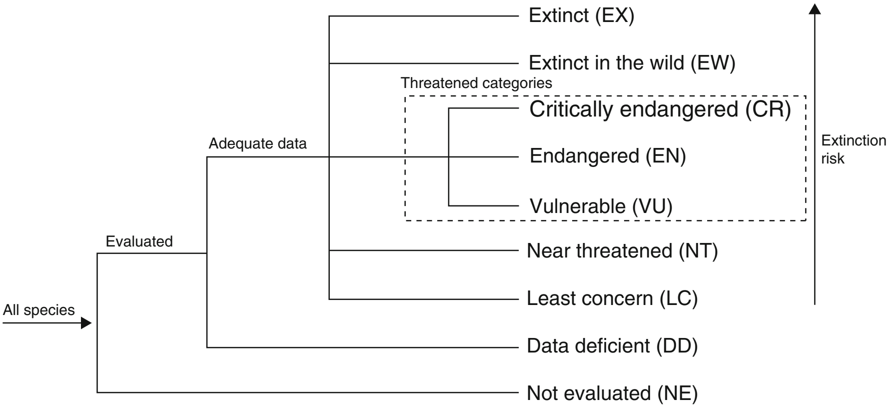


Fig. 14.1 The IUCN red list of threatened species™ provides taxonomic, conservation status, and distribution information on species that have been globally evaluated using the IUCN red list categories and criteria

of the world's overall species [39]. Recent efforts to assess whether vertebrates trends and patterns can be used to extrapolate for other taxonomic groups are ongoing and show promise, with initial assessments indicating that the threat levels seen in invertebrate groups are not dissimilar from that of vertebrates [40].

The IUCN Red List of Threatened Species is the most widely accepted standard for assessing species' risk of extinction [41]. The IUCN Red List assesses species according to five quantitative criteria and classifies them into one of eight categories (see Fig. 14.1). Regular assessments of taxonomic groups provide information on trends in extinction risk, known as the IUCN Red List Index (RLI). The RLI considers species classified as Critically Endangered, Endangered, or Vulnerable to be threatened with extinction and explicitly accounts for Data Deficient species. Comprehensive datasets exist for mammals [42], birds [43, 44], and amphibians [45], while sampled assessments can be used for other taxa [38] (Fig. 14.2).

Trends in vertebrate populations have been measured using the Living Planet Index, a global biodiversity indicator that tracks changes in vertebrate populations in the wild. It contains data on species trends from 1970 to 2005 and is calculated by aggregating population time series for each taxonomic group where data is available [46]. However, results for individual classes are at this point considered preliminary and conservative due to limited data or data biased toward temperate regions [38].

Mammals

Mammals are found across the world's terrestrial and aquatic ecosystems, with South American, sub-Saharan Africa, and Southeast Asia containing the highest concentrations of species of terrestrial mammals. The highest species richness of

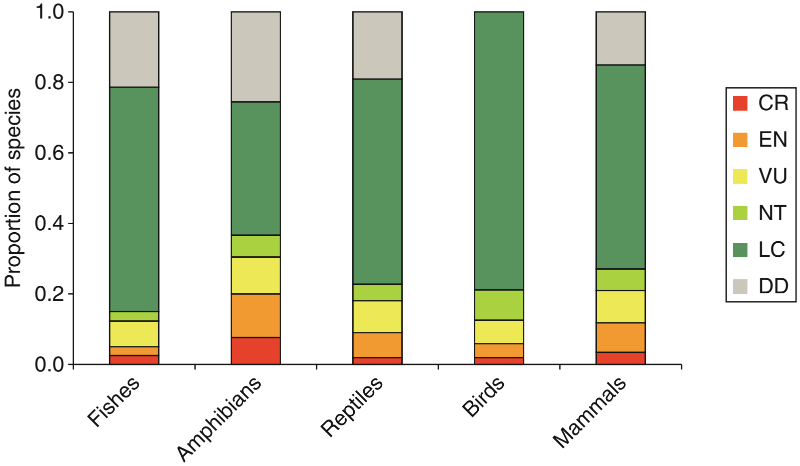


Fig. 14.2 Vertebrate species by IUCN red list categories [38]

terrestrial mammals is found in the tropical rainforests of Brazil and Indonesia and for aquatic mammals, the continental coastlines are the areas of the highest species richness; 21–36% of mammal species are currently threatened with extinction and the highest proportion of threatened species are found in the Monotremes (egg-laying mammals) and Perissodactyla (odd-toed ungulates). Countries with high numbers of threatened terrestrial mammal species include Mexico, Indonesia, Brazil, Papua New Guinea, and Vietnam. For marine mammals, the highest species richness of threatened species is found around the coastal regions of Asia, Japan, North Atlantic, and North Pacific [38, 42].

The LPI shows an overall reduction of 25% for terrestrial mammals. Marine mammals are omitted in the analysis due to sparse data; however, cetaceans have experienced a well-documented decline [47]. Mammal species are threatened from deforestation and coastal development, especially logging, wood harvesting, and smallholder farming. Overexploitation, particularly in the form of hunting for sport, subsistence, or use in traditional medicines, is also a threat to numerous species [38, 42].

Birds

Birds are found on all seven continents, with many migrating thousands of kilometers every year between breeding and wintering grounds. Habitats with high number of bird species include forests, shrubland, grassland, savanna, and inland wetlands [44, 48]. The Neotropics – specifically Colombia, Peru, Brazil and Ecuador – hold the highest numbers of bird species, followed by the Afrotropical, Indomalayan, and Australasian realms. Between 12% and 13% of bird species are currently threatened with extinction; areas with particularly high densities of threatened bird species are the tropical Andes, the Atlantic forests of Brazil, the Eastern Himalayas, Eastern Madagascar, and the archipelagos of Southeast Asia [38, 44].

The LPI shows an overall decline of 8% for birds and 28% for tropical birds alone. The main threats to birds are the spread of logging, wood harvesting, and agriculture in addition to the impacts of invasive alien species. Residential and commercial development, hunting, and pollution are also having serious impacts on bird populations globally [38, 44].

Amphibians

As amphibians are dependent on water for reproduction and other stages of their life cycles, the majority are found in tropical moist forests or freshwater ecosystems. Amphibian species richness is highest in the Neotropical regions of South and Central America and the countries with the highest species richness are Brazil, Colombia, Ecuador, and Peru. The Southeastern USA also has high species richness due to the large number of native salamander species. Roughly 30–56% of all amphibian species are currently threatened with extinction [45]. Among vertebrates, they have the highest proportion of Critically Endangered species and highest number of recent extinctions. The highest species richness of threatened amphibians is found in the dense tropical and subtropical forests of the Northern Andes, Caribbean, Western Ghats, Malaysian Borneo, and West Africa [38, 45].

There is only limited data on amphibian population trends, so the LPI result of an 80% decline should be treated with caution [38]. There is, however, a great deal of supporting evidence that amphibian species have suffered a major decline since the 1970s [40, 45, 49, 50]. Their aquatic habitats are under great human pressures, which included rapid deforestation, habitat degradation, pollution, and overuse of water. In addition, the emergence of the chytrid fungus, *Batrachochytrium dendrobatidis*, is another major threat. First described in the 1990s, the fungus is now found on every continent inhabited by amphibians and evidence suggests that it is responsible for causing mass mortalities among many populations [51, 52].

Fishes

The West and Central Indo-Pacific, South China, and Coral Sea are the areas of the highest marine fish species richness, while Southeast Asia, Eastern Africa's Great Lakes, South American Amazonia, and the forest of Central Africa have the highest species richness of freshwater fish [38]. Of the world's fish species, 41% are obligate freshwater, 58% are obligate marine, and the remaining 1% can tolerate both systems. An assessment of a representative sample of fish species suggests that between 12% and 34% are currently threatened with extinction. The highest species richness of threatened marine fishes is found in the Indo-West Pacific, Central Indo-Pacific, South China, Coral, Caribbean, and Mediterranean Seas, and the highest species richness of threatened freshwater fish species is found in Southeast Asia [38].

The available data for the LPI for freshwater fish indicates a rapid and steady decline in population abundance approaching a 65% reduction. Marine data are limited, but indicate a gradual decrease of 20% in the number of marine fish populations [38]. The main threat to fish species is overexploitation and pollution; aquatic habitats are routinely treated as limitless sources for human consumption [53, 54]. Over 85% of the world's fisheries are either recovering or fully or partially overexploited [55]. In addition, freshwater fish have been particularly negatively affected by pollution and habitat alteration caused by damming and water management activities such as abstraction for agriculture [56].

Reptiles

While reptiles are the dominant vertebrate of arid systems, like other vertebrates, the highest species richness is found in tropical forests. Indonesia and the Congo Basin have the highest species richness of reptiles. Marine reptiles, such as sea snakes, turtles, and crocodiles, are found in the highest concentrations in the West and Central Indo-Pacific and the South China and Coral Seas. An assessment of a representative sample of species suggests that between 18% and 32% of reptile species are threatened with extinction, although this varies greatly between groups [38]. For example, a high proportion of crocodilians are threatened compared with a relatively lower number of snakes [57]. The highest species richness of threatened reptiles is found in Southeast Asia and the Ganges basin [38].

The available trend data for the LPI for reptile species indicates that the number of reptiles has declined by 7% [38]. However, this figure for the whole class may mask the decline in certain orders such as turtles and tortoises, which have seen large declines in regional-scale analysis [57]. Habitat loss is the greatest threat to reptiles and is principally in the form of agricultural expansion, logging, and urban development. Hunting, trapping, and overharvesting for consumption and the pet trade are particular threats, especially to turtles and tortoises [58] (Fig. 14.3).

Link Between Species and Ecosystem Function

A key topic in ecology is investigating how the diversity and composition of species in an ecosystem is related to its function. Considerable research has gone into answering this question [59] and ecological theory supports the notion that ecosystem function, and the resultant goods and services, depends on biodiversity. It is

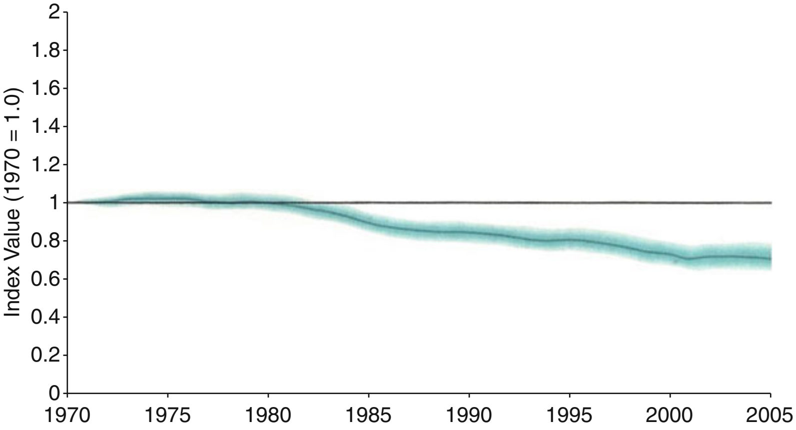


Fig. 14.3 Living planet index for vertebrates. Index value shown in bold; shaded area shows 95% confidence limits [38]

generally predicted that a decrease in biodiversity will result in a reduction of ecosystem function, which will in turn lead to the loss of ecosystem goods and services [60–62]. While there is still work to be done to describe the exact nature of the relationship [63], evidence suggests that species richness, functional characteristics, composition, and interactions influence ecosystem function [64].

The greatest amount of research in this area has focused on the relationship between species richness and ecosystem function [65], with the relationship usually reported as being positive [64]. Evidence suggests that species-rich communities have greater interspecific variation in response to changes and provide more temporally stable ecosystem services [66]. In experimental systems, it has been proven that productivity, nutrient retention, and resistance to invasion and diseases, all tend to increase with increasing species richness [67].

Interactions among species, such as competition, mutualism, and predation, are also crucial to ecosystem functioning [67] and may be more important than species richness [68]. Direct interactions between plants and fungi, plants, and animals, and indirect interactions involving more than two species are essential for ecosystem processes such as transfer of pollen and seeds, transfer of plant biomass production to decomposers or herbivores, construction of habitat complexity, or the spread or suppression of plant, animal, and human pathogens [67]. Interactions between different species trophic levels are also key to regulating the provision of ecosystem services [69]. Certain combinations of species are complementary in their patterns of resource use and can increase average rates of productivity and nutrient retention [64].

Species' functional characteristics also strongly influence ecosystem properties. The loss of species such as ecological engineers or keystone species with unique functional characteristics may have profound effects on ecosystem function [71]. Species composition also plays an important role in ecosystem processes and extinctions of local populations, or their reduction to functionally extinct, can dramatically affect regulating and supporting ecosystem services [71].

Species Loss and Extinction

It is clear that as more species and/or the variability within species disappears, significant loss of ecosystem function, and hence ecosystem goods and services, may occur. While the evolution of species and the extinction of others is a natural process, current extinction rates are estimated to be between 100 and 1,000 times greater than they were in the distant past and future rates may be 10–100 times what they are today [10, 34]. Even among lesser threatened taxa, for example, birds, current extinction rates are several orders of magnitude higher than the natural or background rate. Roughly one-fifth of the world's vertebrates and plants are threatened with extinction [38, 72], and in addition to individual species, entire lineages or ecosystems are threatened [45, 73]. Current rates of population extirpation are probably at least three orders of magnitude higher than species extinction rates [74]. In addition, changes in the relative abundance among species are also capable of causing important changes in ecosystem function [75].

The main cause of extinction across all vertebrate groups is habitat loss or degradation – primarily driven by agricultural development and logging – followed by invasive species and human overexploitation [38, 72]. However as the impacts of climate change increase, it will likely be the greatest driver of extinctions this century. Climate change has already been proven as a cause of species extinction [76] and may likely be the greatest driver of extinction this century. As more distributions shift, it is predicted based on midrange climate warming scenarios for 2050 that 15–37% of species will be committed to extinction, with the most severe impacts resulting from synergistic interactions among a range of threats [77, 78].

Conservation of Species and Ecosystems

Invasive species and human-caused extinctions are already altering ecosystems; there are few places left on the planet that have not been affected by human impacts [79, 80]. Even minor losses may reduce the capacity of ecosystems to adjust to changing environments. As it is often difficult, expensive, or impossible to fix or reverse these changes with technological solutions [64, 81], it is crucial that the diversity of our planet's species and ecosystems is conserved. There are several levels of approach and the appropriate one will depend on the specific goals and resources available. The species level has been the traditional focus of conservation planning and action. Common frameworks include IUCN Species Action Plans and Species Conservation Strategies, which focus on assessing species and making conservation recommendations, and Range-wide Priority Setting, which is designed for widely distributed species. Other approaches include the EDGE approach, which combines IUCN Red List data on threats with species evolutionary distinctiveness to focus on conserving threatened species with few remaining relatives [82].

Conservation approaches at other levels, such as landscape/ecosystems, are also widely used. Examples include Conservation Action Planning [83], which often focuses on a target landscape, and the Ecosystem Approach, which puts natural resource use at the center of decision making [84]. However, even at the ecosystem/ landscape level, conservationists still use species, often iconic and charismatic species ones referred to as flagship species, to act as ambassadors for conservation of the wider area. These species are also a good way to begin engaging with the public, an important part of modern conservation. If people feel a connection to the natural world, they will be more willing to conserve it [85].

Whatever the chosen approach, it must be based on sound science and have a strong conceptual framework to guide action on the ground. Included in this framework should be significant capacity-building component to enable long-term continuation of responses capable of keeping pace with increasing threats, especially climate change. While there have been past conservation successes [73], the conservation community needs to focus more on coordinating large-scale strategic conservation plans around specific objectives. National Biodiversity Strategies and Action Plans, required by Convention on Biological Diversity member countries, have the potential to be very effective, if technical capacity and financial resources are developed to support their creation and implementation [90].

Conclusions

While there is disagreement about the definition and number of species, it is known for sure that species are the building blocks of ecosystems, and their diversity within and among ecosystems is essential to maintaining its function. Species are under threat from numerous anthropogenic factors including habitat loss, overexploitation, introduced and invasive species, and climate change. Current extinction rates are magnitudes higher than those of the past and no one knows for certain how many species the planet can afford to lose before there will be major implications for humanity. As technological solutions to lost or degraded ecosystem services are often expensive, difficult, or even impossible, there is clearly a need to conserve and sustainably manage our ecosystems and the species within them in order to maintain biodiversity and its practical benefits to humanity. However, the current level of conservation action is far outweighed by degree of threat. To halt species loss, there is an urgent need to scale up conservation action. While there have been conservation successes in the past, a coordinated and increased national- and global-level response is urgently need to halt the loss of global biodiversity.

Future Directions

The ecology of species and ecosystems is incredibly complicated and there are still many knowledge gaps regarding species diversity and its exact role within ecosystem function. Further research is needed to more fully explain this relationship and inform conservation action and policy. However, urgent action is also needed to conserve species as it can be confidently predicted that there is already a movement toward the next mass extinction [50, 86]. This is especially critical as the year 2050 is approached when the Earth will need to provide enough natural resources for an estimated 9.2 billion people [87]. Addressing the biodiversity crisis and all its consequences for humanity requires rapid movement toward governance structures and economic systems that encourage the sustainable use of Earth's natural resources. If this effort is to be successful, it cannot be limited to traditional conservationists; it must have the strong support of society as a whole and include leaders from a range of sectors such as engineering, medicine, finance, communications, business, and the arts.

Bibliography

1. Magurran AE, Gill BJ (2011) Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford, UK
2. Hurlbert S (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586
3. Hamilton AJ (2004) Species diversity or biodiversity? J Environ Manage 75:89–92
4. Whittaker RH (1960) Vegetation of the Siskiyou mountains, Oregon and California. Ecol Monogr 30:279–338
5. Mace GM (2004) The role of taxonomy in species conservation. Philos Trans R Soc Lond B Biol Sci 359:711–719
6. Mace GM, Gittleman JL, Purvis A (2003) Preserving the tree of life. Science 300:1707–1709
7. Hey J (2001) The mind of the species problem. Trends Ecol Evol 16:326–329
8. Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge MF et al (eds) Species: the units of biodiversity. Chapman and Hall, London, pp 381–424
9. Mallet J (2006) Species concepts. In: Fox CW, Wolf JB (eds) Evolutionary genetics: concepts and case studies. Oxford University Press, Oxford, UK, pp 367–373
10. Mace GM, Masundire H, Baillie JEM (2005) Biodiversity. In: Hassan R, Sholes R, Ash N (eds) Millennium ecosystem assessment, 2005 Current state and trends: findings of the condition and trends working group Ecosystems and human well-being, vol 1, chapter 4. Island Press, Washington, DC
11. Mayr E (1963) Animal species and evolution. Belknap Press of Harvard University Press, Cambridge, UK
12. Hay J (2000) Genes, categories and species: the evolutionary and cognitive causes of the species problem. Oxford University Press, Oxford, UK
13. Isaac NJB, Mallett J, Mace GM (2004) Taxonomic inflation: its influence on macroecology and conservation. Trends Ecol Evol 19(9):464–469
14. Ward BB (2002) How many species of prokaryotes are there? Proc Natl Acad Sci USA 99:10234–10236

15. Nee S (2003) Unveiling prokaryotic diversity. Trends Ecol Evol 18(2):62–63
16. Mora C, Tittensor DP, Ald S, Simpson AGG, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biol 9(8):e1001127
17. Erwin TL (1982) Tropical forests: their richness in Coleoptera and other Arthropod species. Coleopt Bull 36:74–75
18. Raven PH (1983) The challenge of tropical biology. Bull Entomol Soc Am 29(1):4–12
19. Stork NE (1988) Insect diversity: facts, fiction and speculation. Biol J Linn Soc 35:321–337
20. Stork NE, Gaston KJ (1990) Counting new species one by one. New Sci 1729:43–47
21. Hodkinson ID, Casson D (1991) A lesser predilection for bugs: Hemiptera (Insecta) diversity in tropical rain forests. Biol J Linn Soc 73:101–109
22. Hodkinson ID (1992) Global insect diversity revisited. J Trop Ecol 8:505–508
23. Novotny V, Basset Y, Miller SE et al (2002) Low host specificity of herbivorous insects in a tropical forest. Nature 419:841–844
24. IUCN (2011) IUCN red list of threatened species, version 2011.2: Table 1. www.iucnredlist.org. Downloaded on 14 Nov 2011
25. May RM (2007) Unanswered questions and why they matter. In: May RN, McLean AR (eds) Theoretical ecology: principles and applications, 3rd edn. Oxford University Press, Oxford, UK
26. Rosenzwieg ML (1995) Species diversity in space and time. Cambridge University Press, Cambridge, UK
27. Prendergast JR, Quinn RM, Lawton JH et al (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. Nature 356:335–337
28. Fischer AG (1960) Latitudinal variations in organic diversity. Evolution 14:64–81
29. Hillebrand H (2004) On the generality of the latitudinal diversity gradient. Am Nat 163:192–211
30. Reid WV (1998) Biodiversity hotspots. Trends Ecol Evol 13:275–280
31. Snelgrove PVR (1999) Getting to the bottom of marine biodiversity: sedimentary habitats – ocean bottoms are the most widespread habitat on Earth and support high biodiversity and ecosystem services. BioScience 49:129–138
32. Crame JA (2002) Evolution of taxonomic diversity gradients in the marine realm: a comparison of Late Jurassic and recent bivalve faunas. Paleobiology 28:184–207
33. Gaston KJ, Spicer JI (1998) Biodiversity: an introduction. Blackwell Science, Oxford, UK
34. Baillie JEM, Hilton-Taylor C, Stuart SN (2004) 2004 IUCN red list of threatened species. A global species assessment. IUCN, Gland
35. Myers N, Mittermeier RA, Mittermeier GC et al (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858
36. Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10
37. Williams PH, Humphries CJ, Gaston KJ (1994) Centres of seed-plant diversity: the family way. Proc Roy Soc Lond B 256:67–70
38. Baillie JEM, Griffiths J, Turvey ST et al (2010) Evolution lost: status and trends of the world's vertebrates. Zoological Society of London, London, UK
39. Ponder W, Lunney D (1999) The other 99%: the conservation and biodiversity of invertebrates. Transactions of the Royal Zoological Society of New South Wales, Chipping Norton
40. Collen B, Ram M, Dewhurst N et al (2009) Broadening the coverage of biodiversity assessments. In: Vié J-C, Hilton-Taylor C, Stuart SN (eds) Wildlife in a changing world: an analysis of the 2008 review of The IUCN red list of threatened species. IUCN, Gland, pp 67–77
41. Rodrigues ASL, Pilgrim JD, Lamoreux JF et al (2005) The value of the IUCN red list for conservation. Trends Ecol Evol 21:71–76
42. Schipper J, Chanson JS, Chiozza F et al (2008) The status of the world's land and marine mammals: diversity, threat and knowledge. Science 10(322):225–230
43. Butchart SHM et al (2004) Measuring global trends in the status of biodiversity: red list indices for birds. PLoS Biol 2:2294–2304. doi:10.1371/journal.pbio.0020383

44. BirdLife International (2004) Threatened birds of the World 2004. (http://www.birdlife.org/datazone/species/index.html)
45. Stuart SN et al (2004) Status and trends of amphibian declines and extinctions worldwide. Science 306:1783–1786
46. Collen B, Loh J, McRae L et al (2009) Monitoring change in vertebrate abundance: the living planet index. Conserv Biol 23:317–327
47. Christensen LB (2006) Marine mammal populations: reconstruction historical abundances at the global scale. University of British Columbia, Vancouver, Canada
48. BirdLife International (2008) State of the world's birds: indicators for our changing world. BirdLife International, Cambridge, UK
49. Bielby J, Cooper N, Cunningham AA et al (2008) Predicting declines in the world's frogs. Conserv Lett 1:82
50. Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proc Natl Acad Sci 105:11466–11473
51. Blaustein AR, Wake DB (1990) Declining amphibian populations: a global phenomenon? Trends Ecol Evol 5:203–204
52. Daszak P, Cunningham AA, Hyatt AD (2003) Infectious disease and amphibian population declines. Divers Distrib 9(2):141–150
53. Jackson J, Kirby MX, Berger WH et al (2001) Historical overfishing and the recent collapse of coastal ecosystems. Science 293:629–637
54. Thurstan RH, Brockington S, Roberts CM (2010) The effects of 118 years of industrial fishing on UK bottom trawl fisheries. Nat Commun 1. doi:10.1038/ncomms1013
55. Food and Agriculture Organization of the United Nations (FAO) (2010) The state of the world fisheries and aquaculture. FAO, Rome
56. Le Quesne T, Matthews JH, Von der Heyden C et al (2010) Flowing forward: freshwater ecosystem adaption to climate change in water resources management and biodiversity conservation. World Bank and WWF
57. Reading CJ, Luiselli LM, Akani GC et al (2010) Are snake populations in widespread decline? Biol Lett. doi:10.1098/rsbl.2010.0373
58. Whitfield Gibbon J, Scott DE, Ryan TJ et al (2000) The global decline of reptiles, Déjà vu amphibians. Bioscience 50(8):653–666
59. Naeem S, Wright JP (2003) Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. Ecol Lett 6:567–579
60. Naeem S, Thompson LJ, Lawler SP et al (1994) Declining biodiversity can alter the performance of ecosystems. Nature 368:734–737
61. Daily GC, Alexander S, Ehrlich PR (1997) Ecosystem services: benefits supplied to human societies by natural ecosystems. Issues Ecol 2:1–16
62. Chapin FS, Zavaletta ES, Eviner VT et al (2000) Consequences of changing biodiversity. Nature 405:234–242
63. Luck GW, Daily GC, Ehrlich PR (2003) Population diversity and ecosystem services. Trends Ecol Evol 18:331–336
64. Hooper DU et al (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecol Monogr 75:3–35
65. Kremen C (2005) Managing ecosystem services: what do we need to know about their ecology? Ecol Lett 8:468–479
66. Tilman D (1996) Biodiversity: population versus ecosystem stability. Ecology 77:350–363
67. Diaz S, Tilman D, Fargione (2005) Biodiversity regulation of ecosystem services. In: Hassan R, Sholes R, Ash N (eds) Millennium ecosystem assessment, 2005 current state and trends: findings of the condition and trends working group ecosystems and human well-being, vol 1, chapter 11. Island Press, Washington, DC
68. Grime JP (1997) Biodiversity and ecosystem function: the debate deepens. Science 277:1260–1261

69. Dobson A, Lodge D, Alder J et al (2006) Habitat loss, trophic collapse and the decline of ecosystem services. Ecology 87(8):1915–1924
70. Jones CG, Lawton JH, Shachak M (1994) Organisms as ecosystem engineers. OIKOS 69:3
71. Schindler DE, Hilborn R, Chasco B et al (2010) Population diversity and the portfolio effect in an exploited species. Nature 465:609–612
72. Hoffman M, Hilton-Taylor C, Angulo A et al (2010) The impact of conservation on the status of the world's vertebrates. Science 330:1503
73. Harding S, Clark E, Gardiner-Smith B et al (2010) GLOBE action plan for coral reefs. GLOBE International, London
74. Hughes JB, Daily GW, Ehrlich PR (1997) Population diversity; its extent and extinction. Science 278:689–692
75. Valéry L, Fritz H, Lefueve J-C, Simberloff D (2009) Ecosystem-level consequences of invasion by native species as a way to investigate relationships between evenness and ecosystem function. Biol Invasions 11(3):609–617
76. Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42
77. Thomas CD, Cameron A, Green RE et al (2004) Extinction risk from climate change. Nature 427:145–148
78. Stork NE (2009) Re-assessing current extinction rates. Biodivers Conserv 19(2):357–371
79. McMichael AJ, Butler CD, Folke C (2003) New visions for addressing sustainability. Science 302:1919–1920
80. Halpern BS, Walbridge S, Selkoe KA et al (2008) A global map of human impact on marine ecosystems. Science 319:948
81. Palmer M, Bernhardt E, Chornesky E et al (2004) Ecology for a crowded planet. Science 304:1251–1252
82. Issac NJB, Turvey ST, Collen B et al (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. PLoS One 2(3):e296
83. Hamberg SE, Martin AS (2007) Conservation action planning. Innovations in conservation series. Parks in Peril Programme. The Nature Conservancy, Arlington
84. Convention on Biological Diversity (2000) Decision V/6. Ecosystem approach. In: Proceedings of the fifth conference of the parties, Nairobi
85. Saunders CD (2003) The emerging field of conservation psychology. Human Ecol Forum 10(2):137–149
86. Barnosky AD, Matzke N, Tomiya S et al (2011) Has the Earth's sixth mass extinction already arrived? Nature 471:51–57
87. United Nations, Department of Economic and Social Affairs, Population Division (2007) World population prospects: 2006 revision, highlights. Working Paper No. ESA/P/WP.202
88. Anderson S (1994) Area and endemism. Q Rev Biol 69:451–471
89. Brown JH, Whitman TG, Ernest SKM, Gehring CA (2001) Complex species interactions and the dynamics of ecological systems: long term experiments. Science 293:643–649
90. Baillie JEM, Raffaelli D, Sillero-Zubiri C (in prep) Levels of approach. In: Macdonald D (ed) Key topics in conservation biology 2nd ed. Wiley-Blackwell, Oxford, UK

Chapter 15
Urban Ecology

S.T.A. Pickett and M.L. Cadenasso

Glossary

City	A dense, demographically and economically heterogeneous settlement containing businesses, residences, intensive transportation infrastructure, multistory buildings, and warehousing or manufacturing. Economies of cities focus on services, processing commodities, manufacturing, or finance rather than on agricultural or resource management. Cities are often characterized by a population representing diverse social groups and economic classes in relatively close spatial proximity. As a caution, it should be noted that the term “city” is sometimes used to refer to an entire urban (definition 1) area, and sometimes used in a narrow sense to contrast with suburb (definition 2).
CSE	City-suburban-exurban system. A comprehensive term referring to all components of a complex urban system, metropolitan area, or other cluster of urban areas.
Ecology	The science of the relationship of organisms to each other and the physical environment, and the transformation of resources mediated by those interactions. Human ecology includes the social and institutional structures established by people as components of the system studied.

This chapter was originally published as part of the Encyclopedia of Sustainability Science and Technology edited by Robert A. Meyers. DOI:10.1007/978-1-4419-0851-3

S. Pickett (✉)
Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA
e-mail: picketts@caryinstitute.org

M.L. Cadenasso
Department of Plant Sciences, University of California, One Shields Avenue,
Davis, CA 95616, USA
e-mail: mlcadenasso@ucdavis.edu

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8_15,

Ecosystem	A system comprising the organisms, the physical environment, and the interactions among them within a specified volume of the Earth. Ecosystems may be of any size, depending on the research questions of interest, and are open to material and energetic flows with adjacent systems. Although this definition encompasses human ecosystems, for completeness it is considered that human ecosystems include biological components, the physical environment of air, water, soil, energy, the social and human institutions, and the built environment.
Suburb	This term has two meanings, depending on whether an Old World or a New World context is intended. In the New World, a suburb is a component of a broadly recognized urban area, a primarily residential land cover in which single or multiple household dwellings are interspersed with the open spaces of lawns and generous street landscaping. New World suburbs are often a locus of wealth and power. Modest commercial nodes, mostly as service rather than manufacturing or warehousing, may be included in such suburbs. In contrast, Old World suburbs may be less green and more densely built than New World suburbs, and in many cases may host concentrations of lower income, less empowered persons. In the expanding cities of the global south, shanty towns and informal settlements may constitute much of the suburban realm.
Urban	Definition 1 refers to all components of densely settled and built up areas, and contrasts with rural, agricultural, or wild lands. An alternative terminology for this inclusive definition is city-suburban-exurban (CSE) system. Definition 2 refers to dense commercial, industrial, and residential lands in contrast to suburbs and exurbs.

Definition of the Subject

Within the science of ecology, urban ecology is defined as the study of structure, dynamics, and processes in urban ecological systems. Urban ecology is the study of the relationships of human and nonhuman organisms in urban areas, the interactions of these organisms with the native and built physical environment, and the effects of these relationships on the fluxes of energy, materials, and information within individual urban systems and between urban and nonurban systems. Urban ecology applies the methods and concepts of the biological science of ecology to urban areas, but requires and integrates with the concerns, concepts, and approaches of social sciences to produce a hybrid discipline. Urban ecological systems include individual organisms, populations, communities, and landscapes, as well as buildings and infrastructure. Urban ecology further recognizes specific urban ecosystems as a part of the global biogeochemical, economic, and human demographic system.

Importance of Urban Ecology

Urban ecology is important because it brings the insights and knowledge from contemporary biological ecology to bear on urban areas [1]. It replaces the earlier and superseded versions of ecological science that had been used by social scientists, geographers, and urban planners in justifying and predicting urban dynamics in the second half of the twentieth century. Urban ecology as a branch of contemporary ecological science now emphasizes spatial heterogeneity, feedbacks between natural and human system components, probabilistic system change, and the integration between human perceptions and environmental processes.

Urban ecology is also important because urban habitats are increasing worldwide. The United Nations estimates that more than 50% of the global population now resides in urban areas, as defined by the various member nations. In addition, the next three billion people to be added to the world population are expected to live in urban areas. Hence, urban systems are becoming the predominant habitat of humanity, and are an increasingly widespread land cover type worldwide. In the USA, constructed surfaces now cover an area equivalent to that of the state of Ohio [2].

If the disciplines and practices of urban planning and design, ecological restoration, and ecosystem management are to draw more effectively upon ecological knowledge and data, then the science of urban ecology will become an increasingly key resource for these pursuits.

Brief History

Urban ecology has emerged as a subdiscipline of biological ecology only in the last 30 years [3]. It began as an ecological science in the study of the species and biotic communities of conspicuously green patches in cities and metropolises. Parks, vacant lots, disturbed, and derelict lands were the first focal areas of the discipline [4]. More recently, ecologists began to examine areas actively inhabited and managed by people, including lawns and streetscapes [5]. Another contrasting tradition in urban ecology focuses on the coarser scale, to quantify energy and material budgets of cities. This focus, sometimes called urban metabolism, deals with the energy, matter, and information that flow through and are transformed by cities. In all cases, how the biological components and fluxes affect the well-being of people in the city is a concern. However, the contemporary approach to urban ecology differs from the past traditions. First, all areas in the city are now subject to ecological analysis, not just the conspicuous green areas. Second, even in the budgetary approach, the internal spatial configuration of different components of the urban area is recognized as potentially influencing the fluxes and transformations within the larger metropolis. Finally, the fully hybrid nature of the systems is acknowledged, so that cities are seen as neither fully human nor fully natural entities. Rather, they are inextricably both human constructions and biophysical features [6, 7]. Urban ecology was once a study of green spaces in the city. Now it is the study of the ecology of the entire urban area, including biological, built, social, and physical components.

Other scholarly disciplines beyond biology have used the term “urban ecology.” Principal among these has been sociology. This use originated in the 1920s at the University of Chicago under the leadership of Robert Park and Ernest Burgess, who brought concepts of community, competition, and succession that were then current in biological ecology into their new discipline of sociology. Human ecology, which has roots in geography, anthropology, and other social sciences, is closely related to urban ecology when the study subject is urban populations and their interactions. However, other disciplines tend to neglect the physical and biological components of the environment when they address urban ecology.

Introduction

Urban ecology has been used by several disciplines which have different foci and concerns. These contrast or complement the conceptions of urban ecology as a biological science, which is the approach emphasized here.

History

Urban ecology has a long history. The first flowering of urban ecology was a sociological phase established by Park and Burgess at the University of Chicago in the 1920s. Although this was a sociological pursuit, it was centrally informed by analogies from the biological science of ecology, for which the University of Chicago was one of the founding schools. Park, Burgess, and their students explained the unprecedented growth and social change in Chicago in terms of invasion of new human communities, competition among communities, and spatial isolation between different communities and functions in the city. These scholars were disturbed by the doubling of the population of Chicago at the time, and the role of new migrants from the American South or from eastern and southern Europe. The racial, ethnic, and class novelty in the city begged explanation and incited the Chicagoans to seek explanatory and predictive models to serve in the face of such unprecedented changes. This approach to urban ecology was informed by a tacit antiurbanism, as the Chicago sociologists held village and agricultural communities as the paragons of human societies. One of the central tenets of the Chicago school was that cities had a life cycle, analogous to the expected, but incorrect, prediction that ecological communities had predictable life cycles starting from invasion, extending through competition and sorting, and ending in a mature state. This phase of urban ecology ended when social science critics prompted a move toward more individual behavioral explanations of urban change, as opposed to community-based models. A similar, but independent shift occurred in mainstream ecology at about the same time. Even though the academic community moved beyond the deterministic, life-cycle approach to cities, urban policy in the USA continued to assume life-cycle patterns through the 1960s, basing urban conservation and urban renewal policies on this flawed assumption.

Oddly, during the early twentieth century, while their major ideas were informing the birth of sociology and being widely applied in urban systems, most biological ecologists heartily ignored cities and urban systems. European and Japanese ecologists began to explore ecology in urban contexts after World War II. The manifest destruction in the cities in which they lived invited their interest as biologists. What would be the patterns and mechanisms of plant establishment in derelict sites? How would the newly established biotic communities change over time? What benefit might they provide the cities in which they occurred? The questions of the immediate postwar researchers in Europe and Japan were standard ecological questions, but asked in a novel location. This tradition became linked with urban planning in Europe and has remained active in that form [8].

The second wave of urban ecology rose in the 1970s in the USA. Associated with the birth of environmentalism and its concern with the Earth's exponential human population growth, the urban ecology of this era tended to assume that humans were a negative influence on ecosystems, and urban areas provided an extreme case of the human impact that was beginning to worry scientists and the public. A key document from this era is the volume by Stearns and Montag [9]. In it, the problems of urban areas are outlined, and the nature of potential ecologically informed solutions is suggested. However, the ecology of the time was rather coarse-scaled, and assumed equilibrium tendencies of systems, rather than recognizing fine-scale heterogeneity as a causal feature of systems [10]. Furthermore, although failure of the old ecological ideas that had informed the Chicago School was evident, no clear replacement had emerged. Urban ecology in this era concentrated on investigations of conspicuously green patches in the city. Hence, this approach can be characterized as ecology in the city [3]. Parks, cemeteries, gardens, and abandoned lots exemplify this literature.

Another feature of this second wave of urban ecology was a budgetary, systems approach. Epitomized by work in Hong Kong [11], this approach to urban ecology addressed energy and material budgets of cities, and detailed the human costs of pollution and crowding. This approach is characterized as a budgetary feature of ecology of the city. It shares with the early Chicago School an assumption of the importance of urban "pathologies" in the human population. Industrial ecology and urban metabolism are branches from this tradition. Both of these schools of thought analyze the material and energetic inputs, efficiencies, and outputs of urban systems and their components. Life-cycle analysis of materials is a strategy that aims to reduce the use of resources and the generation of wastes associated with contemporary material use. This era of urban ecology did not persist in the USA as a comprehensive field.

A new wave of urban ecology is currently on the rise. It is characterized by several features that differentiate it from prior instances of urban ecology, and make it more comprehensive than earlier approaches. First, it attempts to unify social and biological knowledge, concerns, and approaches [12]. Second, it acknowledges and exploits spatial heterogeneity and fine-scale dynamics as a feature and cause of urban change. Third, it seeks to understand the controls of biogeochemical processes throughout urban systems, including retention, fluxes, and leakage of limiting nutrients and pollutants. Contemporary urban ecology brings the three previously separate goals together for the first time.

Will this current interest in urban ecology wane, as did the previous ones in the USA? One difference between the current manifestation of urban ecology and the previous ones is institutional support. The pioneers of urban ecology in Europe, Japan, and the USA did not have long-lasting research support. As a result, their pioneering efforts were sometimes short-lived. Now there are two urban Long-Term Ecological Research (LTER) sites in the USA, and International Long-Term Ecological Research programs and Zones Ateliers are including urban areas among their rosters. Already the US LTER urban sites are 13 years old. Such longevity promotes interdisciplinary collaboration, continued use of research areas, developing ongoing relationships with communities and decision-makers, and accumulation of lengthy data runs which can expose causal links and the role of pulse events [13]. Acknowledging that urban areas both contribute to and are vulnerable to global changes [13] will tend to keep them in focus in ecological science.

Examples

Urban ecology is such a diverse science that examples are required to give a sense of its breadth.

Patterns of diversity and abundance associated with urbanization are complex and competing explanations exist. Tests of island biogeography theory in urban areas find that species–area relationships are preserved in urban patches [14]. However, in some studies, patch size influenced species composition rather than species richness as a result of organisms at higher trophic levels being preferentially lost from smaller patches [15]. Attempts to directly quantify the extinction and colonization processes that island biogeography relies on have shown that immigration and extinction characterize different kinds of patches [16, 17]. The species composition in a patch is the result of species colonizing the novel habitats formed by urbanization along with those remaining after local extinctions due to isolation or habitat alteration. One prediction of the view of complex causes of urban biodiversity is that urban habitats are not always less diverse than rural patches. Rather, diversity depends on the sum of extinction and colonization rates, which differ regionally and taxonomically. At moderate levels of urbanization, species richness may actually be higher than in nearby wild lands.

A second example is the disconnection between riparian zones of urban streams and the water table [18]. This disconnection limits the capacity of urban riparian zones to convert nitrate, a pollutant in groundwater derived from fertilizer and sewage, to gaseous forms that do not pollute streams. Research in agricultural landscapes has suggested that riparian restoration, inserting woody and grass vegetation between crops and stream banks, is an effective strategy to mitigate nitrate pollution in streams. When the capacity of urban riparian zones to accomplish such mitigation was examined in Baltimore, MD, USA, it was discovered that riparian zones had become disconnected from the groundwater sources that control their ability to convert nitrate to nitrogen gas. With reduced infiltration of stormwater into the ground due to impervious surfaces, and with high incision leaving stranded droughty floodplains in cities, urban riparian zones no longer support the anaerobic conditions and high organic matter required to fuel denitrifying bacteria. Hence, the expected denitrification in urban riparian zones may not always occur [19]. This example demonstrates that knowledge obtained in nonurban environments may not apply to urban situations.

Guide to the Article

It is now clear that urban areas express different combinations of ecological processes than do nonurban areas and that urban areas are an increasingly important component of the global biosphere. It is now possible to examine key principles that are emerging from the new ecological focus on urban research. Future directions for research and management are addressed at the conclusion of the article.

Urban Ecology Principles

Thirteen principles characterize the contemporary science of urban ecology. Of course, such principles are likely to be improved or replaced with advances in this rapidly growing scientific field. The principles can be divided into four groups: (1) the human ecosystem, (2) urban form, (3) urban function, and (4) methodology.

Principles Concerning Human Ecosystems

Much of the history of urban research has proceeded as though city and ecology were different and mutually exclusive. At the same time that the social scientists of the Chicago School were applying bioecological concepts of the day to their work in the cities, another member of the University of Chicago faculty was establishing the concept of plant succession. A member of the Botany Department, Henry

Chandler Cowles, was working on the seemingly pristine plant communities of the Indiana Dunes, distant from Chicago's immigrant-driven hurley burley. There was conspicuously no empirical or theoretical collaboration between the pioneering ecologists and social scientists at the University of Chicago in the early twentieth century [20].

The tradition of treating city and nature as opposites was challenged by adventuresome scholars and practitioners in the closing decades of the twentieth century [21]. This has led to a new conception of cities or urban areas as hybrid socio-bioecological systems. In other words, urban places may be considered to be human ecosystems. As such, they incorporate not only the traditionally recognized biotic and physical components of ecosystems, but also the social structures and built components so conspicuous in cities and towns [7].

The new conception of human ecosystems prompts researchers, planners, and managers in urban systems to study and exploit the reciprocal feedbacks between the social components in all their demographic, institutional, behavioral, and economic complexity, and bioecological processes, whether conspicuous or not. For example, the mitigation of urban heat extremes by trees is well known [22]. Put simply, there is an ecosystem feedback between vegetation and human comfort or risk of heat stress. However, because of the multifaceted nature of the human ecosystem, the embedding of human values and culture, and the linkage of vegetation cover with water demand, the causal link between the desire to mitigate heat stress and the willingness to plant trees is complex. Included are citizen concerns over private property access, commitments to tree maintenance and litter removal on their property, fear of crime associated with vegetated hiding places, risk of treefall, root fouling of infrastructure, and aesthetic judgments [23]. Other examples of such processes are associated with later principles.

Two principles emerge from these considerations about human-natural system coupling in urban contexts:

Principle 1: Cities and urban areas are human ecosystems in which socioeconomic and bioecological processes feed back to one another.

Principle 2: Human values and perceptions are a key link mediating the feedbacks between social and bioecological components of human ecosystems.

Principles Concerning Urban Form

One of the most remarkable features of urban areas is their spatial form. Architects and urban planners have traditionally used figure-ground representations of urban areas, which emphasize the built component of the human ecosystem. Pioneering efforts by landscape architects have shown the need and power of going beyond such classic representations to include both the hidden and the conspicuous bioecological features of urban areas.

Urban form has changed dramatically through time. Early walled cities were physically and socially distinct from the surroundings and often were centered on a cosmologically significant, ceremonial core. The industrial city spread beyond the location of former city walls and established porosity for the purposes of exchange of goods and the accumulation of immigrants needed as labor. After World War II, the industrial city was further disaggregated into a mixture of new suburbs, dispersed commercial districts, and older core cities that were reduced in density. This is the city as part of a network in a dispersed megalopolis [24].

This simple, linear typology does not necessarily reflect the temporal trends in all parts of the world. Indeed in some cities of the global south, no industrial period existed before they began the sprawling growth driven by migration of persons from the provinces seeking opportunity and a better life in the modern city of consumption. Shanty towns outpacing the urban infrastructure are one feature of such cities. Hence, urban form is evolving, which leads to the following principle of urban form:

Principle 3: Urban form is a dynamic phenomenon and exhibits contrasts through time and across regions that express different cultural and economic contexts of urbanization.

Why is urban form important ecologically? The establishment of street grids and major road patterns alters surface drainage [25] and presents barriers and corridors for the movement of native and introduced plants and animals [26]. The adding of new vegetation and the obliteration of the forest, savanna, desert, wetland, or grassland that had previously occupied the urban site follows the dictates and needs of the new urban form. In addition, the infrastructures for supplying clean water and for dealing with fouled water are large alterations to watershed structure and regional water balances. The placement of main sanitary sewers and main storm drains in stream valleys and the shortcutting of ground and surface water flows are components of the universal urban alteration of hydrology. Furthermore, the shapes of heterogeneity introduced by urban form are often rectilinear and of different scale than the former regional and local patterns of environmental gradients and patches. In addition to spatial heterogeneity of the biophysical components of urban systems, the social features of cities, towns, and suburbs are notably patchy [27]. Urbanists have long acknowledged the fine-scale change – often from block to block – in economic activity, wealth, social group, architecture, and land use that characterize cities. Although traditional postwar suburban development typically occupies rather large tracts which they homogenize, the abrupt shifts in urban structure persist. This patchiness is reinforced by the near universal employment of zoning in large urban areas of wealthy countries. All of these heterogeneous structural alterations can have significant effects on urban and adjacent natural ecosystem function, as will be detailed in a later section "Principles Concerning Urban Function". These insights about the nature of urban form at scales as small as residential parcels or a stroll to the corner of the block can be summarized in a principle:

Principle 4: Urban form is heterogeneous on many scales, and fine-scale heterogeneity is especially notable in cities and older suburbs.

The heterogeneous form of contemporary urban systems extends into the surroundings [28]. The suburban fringe in rapidly growing cities may abut farms, desert, forest, or whatever landscape constitutes the predominant rural or wild land cover. Indeed, the adjacencies are often convoluted and complex, leading to interdigitation of urban elements with the wilder or more rural landscape. In some wealthy urban regions, individual homes are embedded in what is otherwise a wild matrix [29]. An example is the building of suburban houses in chaparral shrublands in Mediterranean climates, such as Southern California. A similar case is the construction of houses in the forest fringe of the Pacific Northwest metropolises, or in forested foothills of the exurban lands of the Southern Appalachians in the USA. The insertion of households whose financial equity, lifestyle identity, social connections, and environmental attitudes have been defined by urban life rather than a life of logging, farming, fishing, or other natural resource management is perhaps a more significant interdigitation than the mere presence of their homes. Similar juxtapositions exist for resorts that serve as remote summer or winter destinations for urban dwellers. The interdigitation of urban with wild or less-intensively managed lands results not only from the invasion of wild land by new housing, commercial, or transportation corridors, but also by the afforestation of foothill developments in arid or semiarid climates in which the lower slopes would have been savanna or grassland. The well-to-do suburbs of Oakland, CA, are an example, which now merge to some extent with the ridgetop forests of the Coast Ranges, and where fire is now a real risk. The fact that urban land cover is growing more rapidly than urban population in oil-subsidized economies [24] suggests that such interdigitation will be the source of increasing conflicts and changing exposure to natural disturbance regimes [29]. Recognizing the ongoing interaction of urban and wild lands suggests this principle:

Principle 5: Urban land covers and uses extend into and interdigitate with rural or wild land covers and uses.

Urban form and hence its interaction with less-urbanized landscapes is the result of a complex of causes. One cause of urban form is conformity to regional plans. This cause is rare in many countries, such as the USA, but more common in Europe. Successful constraint on urban sprawl is exemplified in the USA by the metropolitan green line in Portland, OR, or the Urban-Rural Demarcation Line of Baltimore County, MD [30]. In Baltimore County, more than 90% of residents live within the URDL. The line has been in force since the 1970s and is controlled through zoning, the restriction of sewer infrastructure, and the focusing on urban growth in the county to specific corridors. However, the rarity of such regional controls is compensated for by the commonness of the role of developer self-interest in determining urban form. Indeed, urban planners often lament the predominance of the real estate market as a driver of urban form. On the social roster of causes of urban form are those that are described as “push and pull” factors. Persons and households choose to live at specific places in the metropolis based on such pull factors as attraction to open space, commodious housing, affordability, good schools, and access to desirable commercial establishments. Push factors include high taxes, small and old housing, lack of personally controlled open space, perceived risk of crime, underperforming schools, and the like. Firms also experience push and pull factors. In the case of businesses, a customer base, access to efficient delivery routes, free parking, access to a trained work force, and similar concerns are pull factors. The push factors for businesses mirror those experienced by households.

Many of the factors that control urban form are accidental or unintentional. Pollution, traffic congestion, and the concentration of heat in built areas are examples of unintentional disamenities. Urban flooding is an unintended result of successfully draining stormwater from upstream. Drawing down of groundwater is a similar result of successful stormwater drainage. In neighborhoods built on filled wetlands in Boston, such drying of the soil is threatening the wooden piles supporting the foundations of some buildings. These examples suggest attending to both the intended and unintended consequences of urban development and environmental modification. The principle that emerges is the following:

Principle 6: Urban form reflects planning, incidental, and indirect effects of social and environmental decisions.

Because ecology as a science has been absent from cities for so long the need for data is great. However, ecologists cannot rely on the tried and true experimental approach which they employ with such enthusiasm and success in more wild and rural lands. In urban areas, multiple ownerships, the involvement of individuals and communities in land-use decisions even on private land, and the ethical constraints about manipulations that affect people's well-being or pleasure all limit the applicability of controlled, replicated experiments.

However, while ethics and property regimes close one door to experimentation, the processes of urban design, development, and neighborhood revitalization open others [31]. Urban design projects, whether they be landscape architecture projects in the public realm, private development of a subdivision, or installation of environmental retrofits in older residential or commercial areas, can themselves be treated as experiments. Two examples show the power of this approach. One, the classic case of Jordan Cove, CT, compared an alternative suburban development to a traditional layout of homes. The goal was to test the potential to mitigate stormwater quality and amount with minimal effect on real estate development. A second example is the proposed subdivision of a forested tract in the Hudson Highlands region of New York State. The goal in this case was to reduce the impact on a salamander population that is of conservation concern in the State. The landscape architect employed ecological principles to suggest altering the traditional suburban design to reorient roads, preserve small vernal ponds upon which the salamanders depend for breeding, and install rain gardens as a component of the yard of each house to maintain the hydrologic regime of the complex network of ponds and overland drainage. The development is planned to contain drift fences and pitfall traps as infrastructure to continually monitor the salamander population. Hence the development itself can act as an experiment. To work with designs as experiments, partnerships with urban designers and developers will be required. These examples suggest a principle to test in further application:

Principle 7: Urban designs and development projects at various scales can be treated as experiments to expose the ecological effects of different design and management strategies.

Principles Concerning Urban Function

The tradition in both the social and biophysical sciences to see urban areas as distinct and opposed to nature has resulted in a bias against finding natural processes in City-Suburban-Exurban systems. Both the pioneering research in landscape architecture and urban ecology [5] as well as the more recently established projects [13] have, however, confirmed and extended the understanding of ecological processes as parts of cities and their more extensive urban mosaics. Several examples show the power of bioecological phenomena in urban contexts.

In Philadelphia and Boston, buried floodplains, thought to have been engineered in a reliable stormwater management structure or filled to provide substrate for building, continued to manifest higher levels of soil water and were associated with property damage and abandonment [32]. In Philadelphia, the collapse of the sewer that had replaced Mill Creek was a catastrophic outcome of ignoring the function of the urban landscape. Landscape designs envisioned with the participation of neighborhood residents account for hazard and vulnerability, and provide an opportunity to convert a disturbance prone site to an open space amenity. Such sites can contribute constructively to regional stormwater management, recreation, and education.

A second example shows the capacity of open space to perform biogeochemical ecosystem functions in the urban matrix. Nitrate pollution in streams can be used as an index of environmental quality to assess ecosystem function. Researchers who started the Baltimore Ecosystem Study expected suburban lands to be detrimental to the environmental quality of the urban mosaic. This hypothesis was suggested by the large amounts of nitrogen fertilizer and water applied to many American lawns. When small watersheds that drain areas having different amounts of green space in their landscapes were studied however, it was discovered that the suburban subwatersheds exhibited relatively high nitrate retention [18, 33].

A third example of the role of ecological processes is shown in the net carbon flux between surface and atmosphere. In a location near the boundary between Baltimore City and County, an atmospheric flux tower assesses the upward versus downward movement of moisture, temperature, and carbon dioxide, among other factors. Although the urban area is a net generator of carbon dioxide to the atmosphere due to the use of fossil fuels for heating and transportation, there is lower net flux on weekends and when the wind blows across areas of higher vegetation cover compared to winds from areas or greater built cover.

Examples such as these suggest a principle:

Principle 8: Urban areas contain remnant or newly emerging vegetated and stream patches that exhibit bioecological functions.

One of the fluxes that exercises the interest of urban planners and managers worldwide is water. This concern extends well beyond those cities that are located in arid or semiarid environments. Indeed, some cities, such as Phoenix, AZ, and Los Angeles, CA, are so well connected to distant water supplies via diversions from the Colorado River that they may seem impervious to drought. Ironically, although arid-land cities do show concern for water, many cities from moist climates are coming to acknowledge the sensitivity of their reservoir systems to periodic drought [34]. New York City, with its system of impoundments more than 200 km to the north, or Boston with its reservoir in rural central Massachusetts must now plan for the impacts of periodic drought and of the drying that may come with climate change. Even now, development in some counties in eastern Maryland faces limits due to shortages of groundwater, which are the sources for these suburban-style developments.

Such concerns with water supply are not limited to the USA. Urbanization on the Indian subcontinent is sensitive to projected reduced snowpack in the Himalaya Mountains, in which its rivers originate. Water withdrawal for agriculture and forestry in highlands of South Africa stands to alter water availability for which lowland settlement and wildlife compete. Similar diversions along the tributaries of the Nile are of concern for international relationships in East Africa.

Water supply is not the only aspect of water flux that is of concern. All cities, no matter whether they are in moist or dry climates, must deal with periods of high rainfall. Even desert cities are subject to occasional local thunderstorms, or to floods from the foothills and mountains upstream. Hence, stormwater management in spates is as of as much concern in Mediterranean Santiago, Chile, as it is in temperate Seattle, Washington. The traditional mode of stormwater management is to collect it and pass it off downstream as rapidly as possible. While this solves local problems, it creates problems downstream. Stream bed erosion, scouring out the habitats of aquatic organisms and the spawning grounds of anadromous fish, and increasing the temperature of downstream reaches affected by runoff from hot paved surfaces are some of the consequences to the engineering approach to stormwater management. Many urban management plans now reference water supply and stormwater management as key concerns.

A principle emerges from these sorts of events and situations:

Principle 9: The flux of water, including both clean water supply and stormwater management, is of concern to urbanizing areas worldwide, and connects them explicitly to larger regions.

Ecologists and other biologists have recognized the presence of wildlife and plants in cities for a long time. This tradition is especially strong in Europe, but also has early roots in Japan and the USA. Interest has focused both on organisms that represent the native biota and those that represent threats to either native biota or to human health.

Native biological diversity, or biodiversity, is usually reduced in urban areas, with native species declining in richness compared to species introduced from elsewhere [35]. An example of altered species distribution that may have an impact on the future composition of urban forests is the greater number of vine species, especially introduced species, in the treefall gaps found in urban compared to rural forests. Such vines can impede the regeneration of trees in canopy gaps. Still, the total number of species, including both native and exotic species, can be larger in urban areas than in nearby rural areas in the temperate zones. However, in some central German cities, the richness of even native species is greater in urban areas than in the adjacent countryside. Urban wildlife is subsidized by purposeful feeding, by the incidental increase in resources found in garbage, or for fruit-feeding birds, by the large abundance of fleshy-fruited horticultural species planted in settlements [36].

Urban fauna and flora can respond to the novel conditions of urban ecosystems. For example, bird populations in San Diego exhibit genetically different plumage from rural counterparts [37]. Although that difference is not necessarily adaptive, other changes are functional. Urban bird populations have been demonstrated to raise pitch and volume of their songs, presumably selected by the higher noise levels in cities than in the countryside [38]. Plants are well known to genetically adapt to the heavy metals found in brownfields or associated with urban construction.

An important question is the contribution of exotic and native species to ecosystem services in urban areas. Biophilia, the affinity that humans express for natural settings and living things or the salubrious effects of plants and wild animals in the surroundings, are benefits of wildlife and plants in cities. Other services are performed by the biota of cities, including moderation of climate through shading and transpiration, stabilization of soil surfaces and stream banks, and absorption of pollution. These services can be performed by both introduced and native species. Indeed, some introduced species, such as the exotic Norway maple, were imported due to their tolerance of urban stresses. Ironically, this species is one that has escaped into the wild, and due to the deep and seasonally long-lasting shade it casts, poses a threat to the herbaceous species of the forest understory and to regeneration of other canopy tree species [39].

Because of the damage that exotic species can do to native biotic communities and the functional changes they can cause in ecosystems, they are often targeted for removal. Although the introduction, either purposeful or accidental, of new exotic species should be avoided, in urban areas the removal of exotics as a blanket policy may be problematic. The contribution of the species to ecosystem services in urban areas, the ability of native species capable of performing that function to tolerate urban conditions, or the potential to breed native species for these capacities are all issues that must be considered in planning the management of exotic species. Some exotics can contribute to ecological functions in city-suburban-exurban habitats.

A principle that emerges from knowledge about urban biota is this:

Principle 10: Urban biodiversity has multiple components, and the contributions of each kind of species to ecosystem services in cities must be evaluated.

One important aspect of urban function is how equitably the ecological benefits and risks in urban systems are distributed across the human population. This concern, labeled environmental justice, first emerged in communities of color or economically disadvantaged communities that were adjacent to contaminated sites, or sites that produce hazardous pollution. The frequent locations of such communities in topographic positions subject to flooding or landslides are other cases of environmental injustice. An example is the disproportionate damage suffered by low-lying neighborhoods such as the lower 9th Ward of New Orleans during 2005 in the wake of Hurricane Katrina, or the location of informal settlements with their flimsy construction in some South American cities on erodible slopes. Poor and disempowered populations may be actively excluded from desirable sites, as was the case with segregation practices historically in some US cities, or they may be indirectly priced out of locating to sites on higher ground, more distant from polluting businesses, or from facilities that process waste. The patterns and mechanisms of environmental injustice suggest a principle:

Principle 11: Bioecological processes are differentially distributed across the metropolis and the limitation of services and excess of hazards is often associated with the location of populations that are poor, discriminated against, or otherwise disempowered.

Methodological Principles

Two methodological principles of urban ecology complete the survey of guiding concepts for contemporary urban ecological science. These principles draw on conceptual and empirical principles outlined above and can help shape future urban research.

The first methodological principle deals with setting the boundary and scope of the study system. This step is key to quantification and comparison within and across urban systems. However, there is no a priori, universally correct scientific specification of an urban system [40]. Using legally set municipal or town borders may be convenient for policy purposes, but they may not be appropriate limits for scientific research [28]. This is because a large urban system may be made up of many such legally distinct jurisdictions. In addition, materials, energy, and information that are important in structuring an urban system are likely to move across political boundaries. For example, while it is possible and useful to measure species richness in a city, the collection of species so identified may depend on and interact with the larger, regional species pool [14]. A second example compares boundaries chosen for watershed-based studies with those chosen for a policy analysis. The 17,150 ha Gwynns Falls Watershed, which extends from its headwaters in Baltimore County to its confluence with Baltimore Harbor near downtown in the City of Baltimore, provides a sampling transect. Measurements of stream water quality and flow integrate across the two jurisdictions [18]. In contrast, measurements intended to help managers in Baltimore City define tree planting goals and identify priority areas for planting were conducted solely within the city limits. It is the responsibility of the investigators to state what the boundaries of their study units are and what the criteria for inclusion in the urban system are [40]. Such considerations suggest a methodological principle for urban ecology:

Principle 12: Definition of the boundaries and content of an urban study system is set by the researchers based on their research questions or the spatial scope of its intended application.

The second methodological concern is how to frame comparisons within urban systems. Because urban systems are hybrid human–natural complexes and because they are heterogeneous, there are many possible axes of comparison within them [27]. Comparisons may be based on criteria as divergent as the proportion of impervious surface, the amount of woody vegetation, the density of the human population, the surface area of buildings, the accumulated wealth of residents, the lifestyle characteristics of neighborhoods, or the power relationships of local institutions. Many other criteria are possible. However, different criteria are appropriate bases for different comparisons. For example, hydrology and flood risk may be compared on the basis of impervious surface [25]. The amount of woody vegetation may be used in comparisons of the exchange of heat or carbon dioxide between the surface and the atmosphere. The density of human population may be used as a predictor of water quality. Lifestyle contrasts may form the basis for comparisons of environmental decisions made by households in different parts of a metropolis, or in different metropolitan areas [41].

The fact that these many criteria are often patchily arranged suggests that comparisons will rarely be accomplished by running linear transects across metropolitan maps. Although raw distance is sometimes useful in making coarse-scale urban comparisons, as it was along the urban-rural transect in metropolitan New York [42], in many cases the spatial heterogeneity as it exists on the ground may better be presented as a rearranged, abstract gradient [43]. An example of a gradient abstracted from an array of plots is the gradient of wealth along which bird biodiversity was distributed in Phoenix. Thus, while linear transects may be appropriate bases for initial comparison, ordinated gradients are in fact the general case which should be used to frame comparisons over different scales. The principle is as follows:

Principle 13: Urban comparisons can be framed as linear transects or as abstract gradients, and the abstract comparisons acknowledge the spatial complexity of urban heterogeneity.

The baker's dozen of principles present above frame the current state of urban ecology. As a biological discipline closely related to social and economic sciences and the theory and practice of urban design and planning, it is a relatively new and outward-looking field [13]. Hence, these principles are not the final roster that the field may exhibit. Nor are they in what is likely to be their final form. However, these broad ideas should continue to be useful over the next decade as this field consolidates by accumulating more data on specific urban places through time and by comparing amongst urban areas across the globe. These working principles identify assumptions and suggest hypotheses that can be used to guide future research on urban socio-ecological systems.

Future Directions

Although, as shown above, urban ecology has deep roots as a biological science as well as important parallels in social sciences and the design professions, it remains a young discipline. It is poised for significant growth and increased practical importance. There are three realms in which future directions cluster.

Interdisciplinary Integration for Understanding

Urban areas have been studied by many different disciplines, ranging from sociology, physical and human geography, economics, anthropology, and the more recent offshoots of these classic fields, such as political ecology, political economy, and human ecology. There is a large opportunity for integration across these perspectives. Accomplishing that integration is beyond the scope of this article, but ecological science has an important role to play in promoting integration among the diverse perspectives these disciplines represent. Ecology can play this role because it is preeminently an integrative discipline. It incorporates several contrasting perspectives that together define a goal for a more complete understanding of systems that contain biophysical components and biologically driven processes among their complex structures and functions. Ecology as a science focuses on (1) the interactions of organisms and environment through feedbacks, (2) the interactions of organisms with each other, (3) and the transformations that organisms generate within the environment. It is thus a science of interaction, feedback, and change. These three features are among the most important characteristics of urban systems.

Ecology can be a further stimulus to integration within urban systems because of its growing appreciation and study of the role of humans in ecosystems ranging from remote wilderness to suburbs and downtown districts [12]. This biological science has begun to link effectively with economics, with social sciences, with hydrology, and with atmospheric sciences, for example, in order to contribute to more comprehensive models of inhabited and built landscapes [44]. Yet there is still more work to be done on these various frontiers of integration, and the range of urban systems and ecosystem patches within complex city-suburban-exurban systems is an important arena in which to test hypotheses and models combining the perspectives of the different disciplines. Furthermore, data that integrate different disciplinary perspectives and the attempt to generalize cautiously from these data are important empirical activities.

A useful guide to these integrative approaches is the human ecosystem framework. Introduced originally by the social scientist William R. Burch, Jr., and his colleagues [21], this framework has been modified to reflect more bioecological content and thus to better support the integrative program of urban ecology [44]. The framework is not itself a model, but rather an organized roster of potential causes, mechanisms, and interactions upon which specific models can draw. The human ecosystem framework does take a human-centered perspective on inhabited and managed systems and thus breaks the concept of human ecosystem down into the social system as the primary focus and into the biophysical foundations and sociocultural foundations of that social system (Fig. 15.1). This framework is an excellent abstract picture of any urban ecosystem. The components contained in this causal framework will have specific instances and representations in any urban area.

The biophysical foundations start metaphorically with the earth, air, fire (or energy), and water perspectives that must comprise any environment [45]. However, to prevent neglecting important environmental features that ecologists have come to appreciate in all systems they study, it goes on to include nutrients, which at some concentration often become pollutants, toxic materials and contaminants, soil structure and chemistry, vegetation composition and dynamics through time, and the spatial mosaic that the biophysical and built components jointly define in urban systems. Each of these categories has at least as much component complexity, if not more, than appears in Fig. 15.1. The sociocultural foundations include the resources provided by culture in its material, mental, and spiritual forms. Socioeconomic resources among the sociocultural foundation include information, population, labor, and capital. Capital of course refers not only to financial resources but also to the talents and skills of individual persons and the social capital embodied in networks of human interaction. The social system is the structure that enables and constrains the social interactions that all humans depend upon for their immediate well-being. The social system is characterized by the intuitions, in the broad sense, that humans construct to accomplish the tasks of survival, interaction, health, and control. The social system also establishes order among persons and does so in the form of social identity, formal and informal norms of behavior, and the establishment of social rank hierarchies along several dimensions (Fig. 15.1). The framework identifies cycles as a way of recalling that the components of the social system are not fixed in time or space, but can change with the collective changes in individual and household life cycles, individual and group psychological changes, organizational age as expressed in capacity and flexibility, and the degree of persistence of the various institutional structures in a society. The current intensification of globalization suggests that in many cases local or regional dynamics will be affected by biophysical, economic, and social changes that originate at a distance. For example, the massive migrations from countryside to cities, or from countries in which unemployment is high to different countries where people know or perceive there to be greater opportunities, or the displacements by war and disaster, all suggest looking beyond immediate boundaries for causes of change in city-suburban-exurban systems.

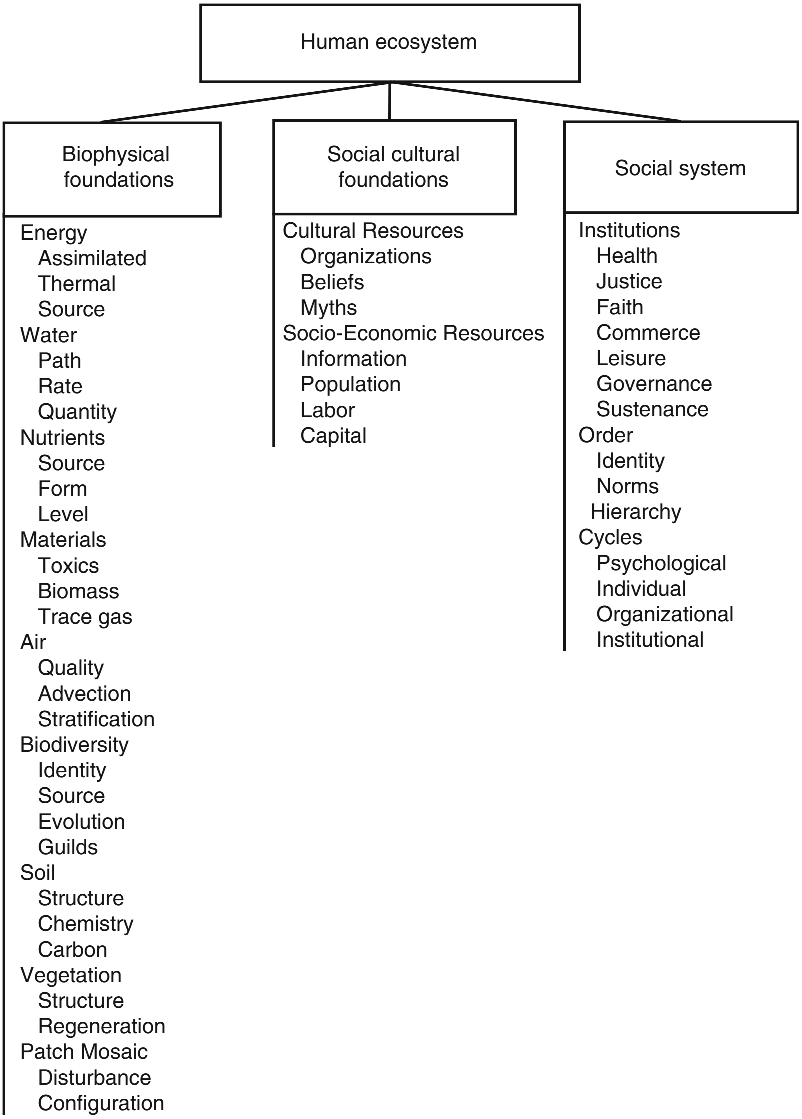


Fig. 15.1 The human ecosystem framework as a hierarchy of potential causes and interactions within human ecosystems (Modified after Machlis et al. [21])

Urban Comparisons

Some generalizations have emerged in discussing the principles. For example, biotic homogenization is a common feature of urban systems [46], although not all taxa or functional groups express homogenization. Likewise, there is an urban stream syndrome, in which downcutting of stream beds is associated with stranding of flood plains and reduction of riparian and in-stream ecological function, and with simplification of the heterogeneity and biotic diversity in stream channels [47]. On a more hypothetical level are generalizations about the increasing similarity of carbon budgets of cities based on planting and maintenance of mesic vegetation in all types of climates [48]. Although these generalizations seem robust so far, they must be challenged, refined, or rejected based on further comparisons with additional cities. Of course, to conduct appropriate comparisons, cities may have to be more explicitly classified according to some globally relevant schema than they are now [27]. City-suburban-exurban systems differ based on location relative to rivers, streams, or coastal waters; regional climate and form of original vegetation; kind of geological substrate; and exposure to natural hazards. In the broad social realm, cities experience different forms of governance; different histories with respect to development, industry, shipping, and ground transportation; access to commodities and sources of wealth; cultural context and diversity; and porosity of social groups. Not all of these may be major axes of comparison, but they point to some of the many dimensions that may affect differences among cities in their bioecological structure and function.

An important aspect of the comparison of cities is their ability to represent global change processes. Cities may well stand in as a laboratory for global change, as temperate cities are now often drier than their mesic surroundings, and due to the urban heat island effect, are generally hotter. Portions of some highly irrigated desert cities may be an exception and in fact be cooler during some hours than the surrounding arid lands due to the evaporation of massive quantities of surface irrigation water or introduced mesic plants [22]. Another feature of global change that some cities may mimic is a great exposure to flood risk due to the building of floodwalls and levees upstream.

Comparisons within cities can benefit from new land cover classifications. The usual “Anderson Level II” classification uses such categories as commercial, industrial, transportation, and residential in low, medium, and high intensities of occupancy. These categories may be too coarse for some desirable ecological comparisons [49]. Likewise, the census geography of block groups based on clusters of approximately 400 households may be rather coarse relative to some levels of ecological function [19]. Both the land use/land cover and the census geographies fail, for example, to adequately match the fine-scale watershed behaviors in urban areas. Social perceptions of edges, enclaves, corridors, and so on may also match other scales of observation than the classical tools of differentiation within urban areas. One example of a classification scheme devised to expose aspects of joint biophysical and social differentiation is the High Ecological Resolution Classification for Urban Landscapes and Environmental Systems (HERCULES) [49]. Rather than assuming land use as the determining criterion, it focuses on land cover in order to provide a structural base to test against bioecological functioning. HERCULES classifies patches in urban systems based on (1) surface characteristics, whether bare or paved, (2) presence and amount of either tall or short vegetation, and (3) cover and kinds of buildings. The classes may be extracted from continuous variation along these three axes, or may be defined to represent percentiles of the different categories. This and other classifications are a frontier tool for urban ecology to promote within and cross-system comparisons.

Integration for Practice

Urban ecological science can be integrated with a number of professional practices. This section illustrates its potential for linkage with urban design and planning, environmental justice, and adaptive processes supporting sustainability. Further integration can emerge from dialog with the concerns of urban policy makers and managers.

Integrating with Urban Design and Planning

Urban design incorporates the activities of architects and landscape architects and often focuses on individual sites, on neighborhoods, or spatially extensive development or redevelopment projects. Urban planning, in contrast, brings aspects of that same knowledge set together with knowledge about policy, governance, and regulation and generates coarse-scale designs for large areas or regions. Master planning is a synonym for this coarser-scale pursuit. Both have traditions that express concern with ecology as a relevant knowledge base or approach. One tradition takes analogies with evolution or ecological processes as a touchstone. This tradition began with Geddes in Scotland, who is one of the founders of modern landscape architecture [50]. Ecological analogies have been common among those who wish to bring an ecological perspective to the urban design and planning professions. This is in part because there has been until recently so little empirical ecological research in urban systems. In this knowledge gap, idealized successional trends from simple to spatially complex systems, ending in a stable "climax," have been brought to bear. Other analogies adopted in the design professions have been the classical ecological expectation of equilibrium in systems. A still larger and deeply flawed analogy has been the adoption of an organism-like life cycle for cities and towns. The life cycle analogy was used to justify policies of physical intervention to circumvent undesired changes in cities. Of course, the assumption that physical change was sufficient to address social problems is another questionable assumption, often tacit, that was used along with life cycle analysis to justify such interventions as slum clearance. Unfortunately, these analogies are not supported by the contemporary knowledge base in ecological science itself. Succession is a probabilistic process that does not lead inexorably toward some ideal state. Equilibrium composition is rarely obtained in ecological systems, even though material balances and thermodynamic processes can be used to understand the structure and function of ecosystems.

In contrast to the use of ecological analogies, which is necessarily limited by the superficiality of the visualizations transferred to design, ecological substance has informed a growing number of designers. In Europe, analyses of "ecotopes," patches that hosted biodiversity or complex bioecological structure, were incorporated into plans to meet aesthetic or functional needs the patches were perceived to satisfy. In the realm of greenfield development, McHarg [51] famously introduced a multilayered strategy to analyze ecological, physical, historical, and cultural features of an area to be developed. The site design was spatially arranged to optimize the joint benefits and reduce the hazards associated with the layers representing different features of the landscape. Spirn [32] applied a careful analytic approach to established urban areas in her work in Boston and west Philadelphia. She used an understanding of watershed dynamics to explain where disempowered communities faced risks due to design in their neighborhood that had been insensitive to ecological processes. This information then informed appropriate neighborhood redevelopment and plans for renewed green space in these old city neighborhoods. This evolving tradition is notable for taking account of both bioecological processes and social concerns. Other designers and planners are promoting the integration of ecological substance with their professional theory and practice [52]. Shared work between designers and ecologists has several advantages. One advantage to working with designers is the entry they provide into the intricacies of development and construction in urban environments, thus making new sites available to their ecological research partners [31]. A second advantage of the increasing interaction between ecologists and designers is to treat constructed design projects as experimental venues [30]. Indeed, because designs provide alternative arrangements of various kinds of surfaces, vegetation, and slopes, for example, they may yield important and novel information on the relationship between urban form and its ecological function [53]. Scaling these relationships from individual small sites to large developments provides another dimension of experimental comparison. Measurements taken in contrasting built projects and landscapes provide an essentially untapped resource for increasing the ecological understanding of urban systems [54].

Environmental Justice

Environmental justice refers to the concern about equal access of all persons to environmental goods and services, and equal avoidance of or capacity to respond to environmental hazards. As a social movement, it emerged from communities of color and from communities with poor access to resources and power, in which people noted an unusual association with toxic sites or other serious environmental disamenities. As a result of this origin, environmental justice has an activist dimension and a scholarly dimension. Ecology can contribute to the scholarship of environmental justice by enhancing the understanding of environmental benefits and environmental hazards. Traditionally, environmental justice has focused on environmental negatives or disamenities. In cities, ecologists and related scientists are increasing the stock of knowledge about not only the disamenities, but increasingly about the benefits that emerge from the biotic components of human ecosystems. This focus on benefits is part of the study of ecosystem services [55].

Examples of urban ecosystem services include the mediation of local climate by trees, the tempering of human behavior by green views, the reduction of stormwater flows, provision of recreation, conversion of some pollutants to neutral forms, and the sequestration of carbon dioxide in biomass, soils, and organic litter [56]. Cultural benefits also accrue to the biotic components of urban ecosystems. These include a sense of place provided by ecological structures such as particular patches of vegetation or streams. Environmental justice can be enhanced by a broader understanding of the spatial distribution of hazards such as fire and exposure to winds, and to the potential for mitigation and other positive services provided by the biotic components of urban ecosystems.

Sustainability and Adaptive Processes

An emerging frontier in urban systems is the nature and processes of sustainability. Sustainability is defined in two main ways. One definition suggests that sustainability is the capacity to meet present needs without compromising the ability of future generations to meet their own needs. A second definition focuses on the reconciliation of demands within economic, social, and environmental spheres that permit the entire integrated system to persist and continue to adapt to change. This second definition is important for giving equal weight to the three facets of the global hybrid system as well as accommodating social justice as a key aspect of the social dimension. Indeed, some scholars define sustainability in terms of economy, environment, and equity, following the assumption that a social system that does not provide for equity is less likely to persist.

Sustainability, as the ability to adjust to changing conditions originating within and beyond the system of interest, is closely related to the concept of resilience. Resilience is the capacity of systems to experience disturbance and shocks and still remain within a given structural or functional domain. Such resilience requires adaptive processes (Fig. 15.2). Adaptive processes are both social [57] and bioecological [58, 59], in keeping with the understanding of most contemporary ecosystems as human ecosystems (Fig. 15.2). Adaptive processes bring the components of human ecosystems into the dynamic realm of sustainability and resilience.

Sustainability has become a goal of many urban jurisdictions. This development is the latest stage in an ongoing evolution of cities in industrial countries. The industrial city was the epitome of overcrowding, contaminated water, and industrially generated pollution, for example. Reformers followed two paths to correct the flaws of the industrial city. One was to abandon it altogether. This strategy motivated the establishment of suburbs, or of garden cities in green sites well outside the shadow of the dark Dickensian industrial city. The second strategy was to alter the industrial city itself. Providing clean water, removing waste, alleviating overcrowding, and reducing the local impact of pollution were hallmarks of this second strategy, which can be labeled, following historian Martin Melosi [60], the “sanitary city.” The sanitary city installed infrastructure

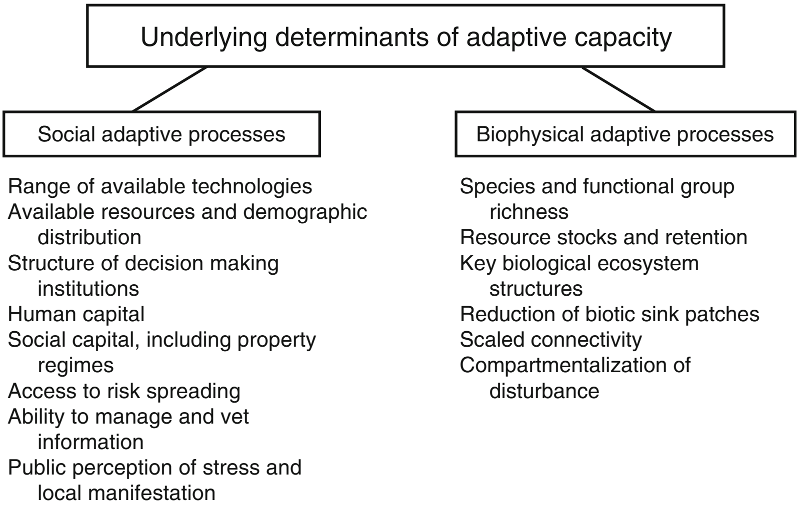


Fig. 15.2 Underlying determinants of adaptive capacity in socio-ecological systems. Social components derived from Yohe and Tol [57] and biophysical components derived from Gunderson and Pritchard [58] and Walker et al. [59]

to carry sewage away from the populated areas. Ultimately, treatment was added, so that downstream systems did not have to bear the burden of the displaced sewage. Smokestacks were built taller so that local pollution was alleviated. Ultimately, newer fuels or scrubbing technologies for coal-fired installations also contributed to reduced pollution. Sanitary landfills were established, and odoriferous wetlands were drained or filled. Many of the solutions employed in the sanitary city were engineered tactics to deal with specific problems. The departmental structure of cities followed this issue-by-issue approach, so that water supply, waste water, solid waste, streets, housing, parks, and so on were dealt with by separate agencies of municipal government.

The emerging sustainable city contrasts with the key characteristics of the older sanitary city. Cities will always be heterotrophic systems, garnering food and other resources from beyond their boundaries. Furthermore, they will generate waste that must be dealt with, requiring commitments of land and other resources – an urban footprint – that extends well beyond their geopolitical boundaries [61]. However, as an ideal, sustainable cities aim to reduce their ecological footprint. A sustainable city is one that reduces its demand and its impact on the environment by reducing resource use and waste generation. Sustainable cities manage potable water, gray water, and stormwater in ways that reduce flood risks, contamination of water, and waste of potable water. They reduce energy use via building standards, landscape design, and reducing per capita use of fuel in transportation. Limiting sprawl while maintaining space for ecological processes and capacities in the urban matrix is an important feature of the sustainable city. Hence, other dimensions of sustainability must be supported. In the environmental realm, reducing vulnerability to natural and human-generated hazards is a sustainability goal, as are supporting native biodiversity and achieving the ecosystem services that eminate from green spaces in the urban matrix. Reduced contribution to global warming can be achieved through enhanced carbon sequestration.

Sustainability by definition also must deal with social processes. Too often in the past, urban interventions have focused on the physical form of the city, assuming that social benefits will follow. Contemporary focus on sustainability recognizes that social capital, community cohesion, and social equity are issues that must be addressed in and of themselves for generating and maintaining liveable cities. Equity is a major component of the social realm, as discussed above, and is reflected in exposure to environmental hazards, access to environmental benefits, and inclusion in the process of making environmentally relevant decisions.

The economic aspect of sustainability is well known, and because of the traditional predominance of attention in urban studies to issues of jobs, finance, the real estate industry, and the relationships of business and political processes, it is not necessesary to review it further in this article. Clearly, however, economic resources and sustainability of economic capacity contribute to well-being in urban systems.

One of the potentially most powerful aspects of the sustainable city ideal is how management and policy are conducted in this new kind of city. As mentioned before, the sanitary city is managed by discrete departments or authorities, often having little cross-communication. For example, the concern with street cleaning are isolated in the transportation department, whereas street sweeping in fact affects stormwater quality. Another cross-cutting issue is how the presence of vegetation can influence aggression, which joins the concerns of a parks department with those of public safety. These simple examples suggest that sustainability can be enhanced by integrated management in urban systems. Indeed, cross-cutting management and policy are hallmarks of the emerging sustainable city. This suggests that sustainability can be usefully considered a central function of city administration and not an isolated pursuit to be marginalized in a special office isolated from the traditional activities of municipal government. Indeed, there are growing numbers of examples of cross-sectoral management as part of sustainability strategies. For example, in the city of Baltimore, neighborhood tree planting and gardening, removal of unneeded pavement in school yards, altered street cleaning schedules, community environmental activities, and installation of fine-scale best management practices such as rain gardens specifically designed to be both attractive and to allow infiltration of stormwater into the soil, engage multiple city departments in specific neighborhoods to achieve social, environmental, and economic revitalization. Ecological research can help inform and evaluate sustainability plans, which are often mostly metaphorical or assessed only by indices of human outcomes.

Communication between researchers and decision-makers is a key ingredient in the success of the sustainable city. Growing experience shows that communication is enhanced when approached as a two-way dialog, rather than a one-way flow of information from science to policy. Indeed, such a dialog can help shape scientific research and identify practical management projects that provide data about the structure and function of urban ecosystems. Mutual respect for the knowledge, constriants, and rewards of each group involved in the dialog is a firm foundation for effectively shaping the science and the practice of urban sustainability. Establishing effective and lasting forums for such communication is a practical part of the frontier of urban ecological science. Such dialog is one tool for the sustainable city, which remains a goal rather than a reality. Combining the knowledge, concerns, and networks represented by problem solving and knowledge generation is perhaps the most urgent frontier for further research and action.

Bibliography

Primary Literature

1. Collins JP et al (2000) New urban ecology. Am Sci 88:416–425
2. Elvidge CD et al (2004) US constructed area approaches the size of Ohio. Euro Opt Soc 85:233–240
3. Grimm NB et al (2000) Integrated approaches to long-term studies of urban ecological systems. BioScience 50:571–584
4. Numata M (1977) The impact of urbanization on vegetation in Japan. In: Miyawaki A, Tuxen R (eds) Vegetation science and environmental protection. Maruzen, Tokyo
5. Loucks OL (1994) Sustainability in urban ecosystems: beyond an object of study. In: Platt RH, Rowntree RA, Muick PC (eds) The ecological city: preserving and restoring urban biodiversity. University of Massachusetts Press, Amherst, pp 48–65
6. Alberti M et al (2003) Integrating humans into ecology: opportunities and challenges for studying urban ecosystems. BioScience 53:1169–1179
7. Pickett STA, Grove JM (2009) Urban ecosystems: what would Tansley do? Urban Ecosyst 12:1–8
8. Sukopp H (1990) Urban ecology and its application in Europe. In: Sukopp H, Hejny S, Kowarik I (eds) Urban ecology: plants and plant communities in urban environments. SPB Academic, The Hague, pp 1–22
9. Stearns F, Montag T (eds) (1974) The urban ecosystem: a holistic approach. Dowden, Hutchinson and Ross, Stroudsburg, p 217
10. Wu J, Loucks OL (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. Q Rev Biol 70:439–466
11. Boyden S et al (1981) The ecology of a city and its people: the case of Hong Kong. Australian National University Press, Canberra, pp 21–473
12. Redman C, Grove JM, Kuby L (2004) Integrating social science into the long-term ecological research (LTER) network: social dimensions of ecological change and ecological dimensions of social change. Ecosystems 7:161–171
13. Grimm NB et al (2008) Global change and the ecology of cities. Science 319:756–760
14. Fernández-Juricic E, Jokimäki J (2001) A habitat island approach to conserving birds in urban landscapes: case studies from southern and northern Europe. Biodivers Conserv 10:2023–2043
15. Gibb H, Hochuli DF (2002) Habitat fragmentation in an urban environment: large and small fragments support different arthropod assemblages. Biol Conserv 106:91–100
16. Crooks KR et al (2001) Extinction and colonization of birds on habitat islands. Conserv Biol 15:159–172
17. Leston LFV, Rodewald AD (2006) Are urban forests ecological traps for understory birds? An examination using northern cardinals. Biol Conserv 131:566–574

18. Groffman PM et al (2003) Down by the riverside: urban riparian ecology. Front Ecol Environ 1:315–321
19. Cadenasso ML et al (2008) Exchanges across land-water-scape boundaries in urban systems: strategies for reducing nitrate pollution. Ann NY Acad Sci 1134:213–232
20. Light JS (2009) The nature of cities: ecological visions and the American urban professions 1920–1960. Johns Hopkins University Press, Baltimore
21. Machlis GE, Force JE, Burch WR (1997) The human ecosystem. 1. The human ecosystem as an organizing concept in ecosystem management. Soc Nat Resour 10(4):347–367
22. Baker LA et al (2002) Urbanization and warming of Phoenix (Arizona, USA): impacts, feedbacks and mitigation. Urban Ecosyst 6:183–203
23. Troy AR et al (2007) Predicting opportunities for greening and patterns of vegetation on private urban lands. Environ Manage 40:394–412
24. Alig RJ, Kline JD, Lichtenstein M (2004) Urbanization on the US landscape: looking ahead in the 21st century. Landscape Urban Plann 69:219–234
25. Gurnell A, Lee M, Souch C (2007) Urban rivers: hydrology, geomorphology, ecology and opportunities for change. Geogr Compass 1(5):1118–1137
26. Snep RPH et al (2006) How peri-urban areas can strengthen animal populations within cities: a modeling approach. Biol Conserv 127:345–355
27. Dow K (2000) Social dimensions of gradients in urban ecosystems. Urban Ecosyst 4:255–275
28. Theobald DM (2004) Placing exurban land-use change in a human modification framework. Front Ecol Environ 2:139–144
29. Radeloff VC et al (2005) The wildland urban interface in the United States. Ecol Appl 15:799–805
30. Outen DC (2007) Pioneer on the frontier of smart growth: the Baltimore County, MD experience, in Smart growth @ 10: a critical analysis of Maryland's landmark landuse program. The National Center for Smart Growth Research and Education, Annapolis and College Park, Maryland, p 49
31. Felson AJ, Pickett STA (2005) Designed experiments: new approaches to studying urban ecosystems. Front Ecol Environ 3:549–556
32. Spirn AW (2005) Restoring Mill Creek: landscape literacy, environmental justice and city planning and design. Landscape Res 30:395–413
33. Groffman PM et al (2002) Soil nitrogen cycling processes in urban riparian zones. Environ Sci Technol 36:4547–4552
34. Kennedy C, Cuddihy J, Engel-Yan J (2007) The changing metabolism of cities. J Ind Ecol 11:43–59
35. McKinney ML (2002) Urbanization, biodiversity, and conservation. BioScience 52:883–891
36. Beissinger SR, Osborne DR (1982) Effects of urbanization on avian community organization. Condor 84:75–83
37. Yeh PJ (2004) Rapid evolution of a sexually selected trait following population establishment in a novel habitat. Evolution 58:166–174
38. Slabbekoorn H, Peet M (2003) Birds sing at a higher pitch in urban noise. Nature 424:267
39. Wyckoff PH, Webb SL (1996) Understory influence of the invasive Norway maple (*Acer platanoides*). Bull Torrey Bot Club 123:197–205
40. McIntyre NE, Knowles-Yanez K, Hope D (2000) Urban ecology as an interdisciplinary field: differences in the use of "urban" between the social and natural sciences. Urban Ecosyst 4:5–24
41. Zhou W et al (2008) Can money buy green: demographic and socioeconomic predictors of lawncare expenditures and lawn greenness in urban residential areas. Soc Nat Resour 22:744–760
42. Pouyat RV, Turechek WW (2001) Short- and long-term effects of site factors on net N-mineralization and nitrification rates along an urban-rural gradient. Urban Ecosyst 5:159–178
43. McDonnell MJ et al (1997) Ecosystem processes along an urban-to-rural gradient. Urban Ecosyst 1:21–36

44. Pickett STA et al (2011) Urban ecological systems: scientific foundations and a decade of progress. J Environ Manag 92:331–362
45. Spirn AW (1984) The granite garden: urban nature and human design. Basic Books, New York, pp 1–334
46. Clergeau P et al (2006) Avifauna homogenisation by urbanisation: analysis at different European latitudes. Biol Conserv 127:336–344
47. Walsh CJ et al (2005) The urban stream syndrome: current knowledge and the search for a cure. J N Am Benthol Soc 24:706–723
48. Carreiro MM et al (2009) Carbon and nitrogen cycling in soils of remnant forests along urban-rural gradients: case studies in the New York metropolitan area and Louisville, Kentucky. In: McDonnell MJ, Hahs A, Breuste J (eds) Ecology of cities and towns: a comparative approach. Cambridge University Press, New York, pp 308–328
49. Cadenasso ML, Pickett STA, Schwarz K (2007) Spatial heterogeneity in urban ecosystems: reconceptualizing land cover and a framework for classification. Front Ecol Environ 5:80–88
50. Welter VM (2002) Biopolis: Patrick Geddes and the city of life. MIT Press, Cambridge, p 355
51. McHarg I (1969) Design with nature. Doubleday/Natural History Press, Garden City, p 197
52. Musacchio LR (2009) Pattern:process metaphors for metropolitan landscapes. In: McDonnell MJ, Hahs A, Breuste J (eds) Ecology of cities and towns: a comparative approach. Cambridge University Press, New York, pp 484–502
53. Pickett STA, Cadenasso ML (2008) Linking ecological and built components of urban mosaics: an open cycle of ecological design. J Ecol 96:8–12
54. Cook WM et al (2004) Learning to roll with the punches: adaptive experimentation in human dominated systems. Front Ecol Environ 2:467–474
55. Boone CG et al (2009) Parks and people: an environmental justice inquiry in Baltimore, Maryland. Ann Assoc Am Geogr 99:767–787
56. Bolund P, Hunhammar S (1999) Ecosystem services in urban areas. Ecol Econ 29:293–301
57. Yohe G, Tol RSJ (2002) Indicators for social and economic coping capacity – moving toward a working definition of adaptive capacity. Glob Environ Change 12:25–40
58. Gunderson LH, Pritchard L Jr (eds) (2002) Resilience and the behavior of large-scale systems. Island Press, Washington, DC, p 287
59. Walker B et al (2004) Resilience, adaptability and transformability in social-ecological systems. Ecol Soc 9(2): Article 5
60. Melosi MV (2000) The sanitary city: urban infrastructure in America from colonial times to the present. Johns Hopkins University Press, Baltimore, 600 pp
61. Rees WE (1992) Ecological footprints and appropriated carrying capacity: what urban economics leaves out. Environ Urbanization 4:121–129

Books and Reviews

Alberti M (2008) Advances in urban ecology: integrating humans and ecological processes in urban ecosystems. Springer, New York

Beatley T (2000) Green urbanism: learning from European cities. Island, Washington, DC

Bullard RD (ed) (2005) The quest for environmental justice: human rights and the politics of pollution. Sierra Club Books, San Francisco

Burch WR Jr (1988) Human ecology and environmental management. In: Agee JK, Johnson DR (eds) Ecosystem management for parks and wilderness. University of Washington Press, Seattle, pp 145–159

Burger J (1999) Animals in towns and cities. Kendall/Hunt Publishing Company, Dubuque

Cadenasso ML, Pickett STA (2008) Urban principles for ecological landscape design and management: scientific fundamentals. Cities Environ 1(2):1–16

Catterall CP (2009) Responses of faunal assemblages to urbanisation: global research paradigms and an avian case study. In: McDonnell MJ, Hahs A, Breuste J (eds) Ecology of cities and towns: a comparative approach. Cambridge University Press, New York, pp 129–155

Curwell S, Deakin M, Symes M (eds) (2005) Sustainable urban development, volume 1: the framework and protocols for environmental assessment. Routledge, New York

Forman RTT (2008) Urban regions: ecology and planning beyond the city. Cambridge University Press, New York

Frumkin H, Frank L, Jackson RB (2004) Urban sprawl and public health: designing, planning and building for healthy communities. Island, Washington, DC

Gleick PH (2009) The world's water: 2008–2009. Island, Washington, DC

Grove JM (2009) Cities: managing densely settled social-ecological systems. In: Chapin FS III, Kofinas GP, Folke C (eds) Principles of ecosystem stewardship: resilience-based natural resource management in a changing world. Springer, New York, pp 281–294

Gunderson LH (2000) Ecological resilience – in theory and application. Annu Rev Ecol Syst 31:425–439

Marzluff JM, Schulenberger E, Endlicher W, Alberti M, Bradley G, Ryan C, Simon U, ZumBrunne C (eds) (2008) Urban ecology: an international perspective on the interaction between humans and nature. Springer, New York

McDonnell MJ, Breuste J, Hahs A (2009) Introduction: scope of the book and need for developing a comparative approach to the ecology of cities and towns. In: McDonnell MJ, Hahs A, Breuste J (eds) Ecology of cities and towns: a comparative approach. Cambridge University Press, New York, pp 1–5

McGrath BP, Marshall V, Cadenasso ML, Grove JM, Pickett STA, Plunz R, Towers J (eds) (2007) Designing patch dynamics. Columbia University Graduate School of Architecture, Preservation and Planning, New York

Sukopp H, Numata M, Huber A (1995) Urban ecology as the basis of urban planning. SPB Academic, The Hague

United Nations Population Fund (2007) State of world population 2007: unleashing the potential of urban growth. United Nations Population Fund, New York

Index

R. Leemans (ed.), *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4614-5755-8,