Field and Theoretical Studies
in Network Ecology

by

Yevgenya Shevtsov

(Under the Direction of Bernard C. Patten)

Abstract

This dissertation consists of a study that examines the importance of indirect energy flows in food webs and a multi-taxon gradient analysis study informed by a network perspective.

Food Webs and Network Analysis

Compartment models are widely used to represent ecological networks of stocks and flows of conserved substances. Network environ analysis (NEA) has revealed several interesting properties of flow–storage networks but can only be applied to systems at a constant steady state. I developed a computational analog of NEA called dynamic environ approximation (DEA), which can be used away from steady state. I used DEA to examine the effects of system size and connectance on the importance of indirect energy flows in a commonly studied theoretical food web model. Over the full range of parameter values examined, the mean fraction of energy traveling over indirect paths was 9.2%, but could be as high as 30%. This quantity increased with system size but peaked at intermediate connectance levels, a pattern explained by the availability of more pathways at intermediate connectance levels.

Multi-Taxon Gradient Analysis

The extent to which ecological communities are coherent entities as opposed to mere intersections of species distributions is one of the fundamental questions of ecology. Gradient analysis is commonly used to address this question; however, all such studies have used organisms from a single guild. This risks missing connections due to non-competitive interactions, which should be most common among functionally different organisms. I used two different methods of analyzing species abundance data, elements of metacommunity structure (EMS) and causal discovery, to examine the importance of species interactions in structuring communities. The EMS analysis found that the distributions of study taxa commonly exhibited high coherence, turnover and boundary clumping, the pattern termed "Clementsian" in EMS. Also, pairs and triplets of directly interacting guilds had higher-than-expected boundary conjunction values, while those that did not directly interact generally did not. I also produced a causal interaction network for my study taxa and found that inter- and intra-guild interactions were equally common. These results highlight the importance of inter-guild interactions in structuring patterns of cooccurrence.

Index words:

Ecological networks, Environ analysis, Food webs, Indirect effects, Ecological communities, Gradient analysis, Elements of metacommunity structure

Field and Theoretical Studies in Network Ecology

by

Yevgenya Shevtsov

B.S., University of California Los Angeles, 2005

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial
Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2012

Field and Theoretical Studies in Network Ecology

by

Yevgenya Shevtsov

Approved:

Major Professors:	Bernard C. Patten
Committee:	Caner Kazanci Jacqueline E. Mohan Chris J. Peterson Amy D. Rosemond

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2012

Dedication

To all my teachers,

whether or not they ever officially had that title.

Acknowledgments

First, I want to thank Rick Vance and Kristin McCully. Their help and encouragement allowed me to go from being a math-phobic undergraduate to the author of the theoretical research described here, and the approach to modeling hammered into me in Rick Vance's class continues to inform my work.

Rosalyn Rael and Neo Martinez provided crucial support (and code) for the food web modeling. My field work depended on assistance from many people. I once asked Jim Porter for help on a grant application and received two hand-written pages of feedback that resulted in my getting the funding. Amy Rosemond and Brian Kloeppel helped arrange facilities at the Coweeta field site. Rebecca Bourquin, Charmaine Brooks, Noelle Ntsama, and Dan and Jenny Williams provided invaluable field and lab assistance. Dave Porter, Marshall Darley, Jim Leebens-Mack, Paul Hendrix and Kamal Gandhi generously allowed the use of their labs at UGA, while Jay Garland, Michelle Birmele, Tina Khodadad and the Kennedy Space Center Space Life Sciences Lab provided access to their molecular lab's expertise and equipment. Steve Presley and Joseph Ramsey advised on data analysis. The work was partly funded by a Sigma Xi Grant-in-aid of Research to Jane Shevtsov.

My committee members, Caner Kazanci, Jackie Mohan, Amy Rosemond and Chris Peterson, as well as Paul Hendrix, who retired before I finished, have helped with everything from planning my research and handling logistics to debugging code. Thank you for your availability and patience.

Alan Garfinkel, in addition to helping me think through some data analysis, has been an extraordinary mentor. When I had to move back to Los Angeles and finish remotely, he provided me with an opportunity to teach and intense, wide-ranging conversations that helped me keep going.

Bernie Patten, my advisor, took a chance on a would-be ecologist in a wheelchair and has been putting up with an argumentative graduate student ever since. The most important thing I learned from him was not network analysis but independence – how to do ecology with little money and even less attention to the conventional metrics of success. This experience will undoubtedly shape my future career.

Finally, my parents, Tatyana and Vladimir Shevtsov, worked and fought for my education since shortly after we arrived in the US. They let me drag them all over the country and were willing to be the field assistants of last resort. After moving back to LA, they never let me forget that finishing this dissertation was priority #1. I thank them for everything.

Preface: Simplicity

The woods are rich.
Acorns and hickories,
toads, chipmunks, monarch butterflies
Glorious autumn days
when beech leaves green the sunlight
and poplars paint the ground

And yet I work with symbols
Bloodless on the page,
but lucid, clear
Hieroglyphs showing verbs,
flows that I glimpse for moments,
when ducks dabble in a lake
or a hawk clutches
the bloody remains of a squirrel

I go
from outside to inside
and then out again
Wanting to transcend

the hard facts of the land,
the details of this time and place

Waiting for the landscape to become transparent
and to unfold
its intricate simplicity
within my mind

Contents

Chapter 1

Introduction

One of the most common definitions of ecology is "the study of how organisms interact with each other and their nonliving environment" (Brown, 1999). These interactions – predation, competition, parasitism, pollination, seed dispersal and a myriad others – form networks of strong and weak links. While a single link may be isolated for study, in the real world, it is always part of a network, affecting and being affected by it.

This dissertation describes network-explicit and network-implicit studies of ecological networks with a focus on trophic interactions. By "network-explicit studies", I mean those that examine the network as a network. The network-explicit studies described here are primarily theoretical in nature; however, inferring a causal network from data, as in Chapter 5, is a network-explicit research. Network-implicit studies, on the other hand, are informed by the idea of an interaction web but do not model this web explicitly. The multi-guild gradient analysis described in Chapter 5 is such a study. It envisions a community consisting of many distinct guilds as a network of interdependencies created by fundamental differences in how organisms make a living, similar to the network of international trade ties created by economic specialization (Agnew, 2002).

This dissertation consists of a literature review, a new computational method for analyzing flows in ecological models, an application of this method to a commonly studied food web model, and a field study of the importance of interspecies interactions in structuring biological communities. Chapter 2, the literature review, examines the different questions that have been asked about ecological networks and how they have been answered. It also provides an overview of environ analysis, a network analysis method that forms the basis for Chapters 3 and 4, and argues in favor of using network properties to explain ecological phenomena.

Traditional environ analysis can only be applied to systems at a constant steady state. Chapter 3 develops a computationally-oriented generalization of environ analysis, termed dynamic environ approximation (DEA), and demonstrates its use with some simple models. Chapter 4 then applies DEA to a commonly studied food web model with the goal of evaluating the importance of indirect energy flows in food webs and finding out what features of food web structure determine this, with a focus on the effects of network size and connectance.

Chapter 5 looks at a classic question in ecology – the extent to which ecological communities are coherent entities as opposed to mere intersections of individual species distributions – from a network point of view. It simultaneously examines the abundances of taxa belonging to four interacting guilds – seed plants, soil basidiomycetes, collembola and mesostigmatid mites – along an elevation gradient, evaluates the coherence of communities, and tests the hypothesis that coherence is positively related to the number of interacting guilds studied. I also use a causal discovery algorithm to construct an interaction network for the taxa studied and investigate whether interguild interactions are more common than intraguild ones. Finally, Chapter 6 reviews the value and limitations of network approaches to ecology.

Chapter 2

Literature Review: Ecological Network Questions and Explanations

"The question is, what is the question?" –John Archibald Wheeler, physicist, 1911-2008

Ecological networks represent systems of pairwise interactions between species or components of ecosystems. These interactions may be of any type – antagonistic (as in food webs and host-parasite networks), mutualistic, competitive, or purely physico-chemical.

Many good reviews of both the current state and historical development of network ecology are available (e.g. Fath and Patten, 1999b, Proulx et al., 2005, Dunne, 2006, Bersier, 2007, Ings et al., 2009), and I will not try to repeat their content here. Rather, this review will focus on the questions that have been asked about ecological networks and the ways in which they have been answered. Such a perspective may reveal areas for future research and clarify or bring into question the assumptions that underlie our research practices. In particular, I will discuss how features of networks have been explained and how they have been used to explain ecosystem and community properties.

The most fundamental type of ecological network research is descriptive. This category includes research that simply seeks to identify what interacts with what, as well as quantitative studies of matter and energy flow in ecological networks. Studies of network properties that can only be seen by analyzing models (e.g. Patten, 1986) are also included in this category if they are oriented toward describing or quantifying these properties more than explaining them or examining their consequences. A related line of research asks how networks and network properties change over time. This kind of work spans the spectrum from mainly descriptive (e.g. Borrett et al., 2006) to mainly explanatory (e.g. Virgo et al., 2006).

Another prominent type of research attempts to explain why ecological networks have the structure that they do. This question has been important in the study of food webs and was later adopted by ecologists studying mutualistic networks; however, it has rarely been asked about ecosystem models except when highly detailed ones have been incorporated into food web studies. This disparity may arise from the fact that many ecosystem models are highly aggregated, so their structure is strongly driven by the modeler's choices. Other studies look at how disturbances, particularly anthropogenic ones, affect network structure and function (e.g. Cross et al., 2007, Miehls et al., 2009), and how a compartment's position in a network affects the compartment's properties (e.g. Patten and Witkamp, 1967).

Finally, some research tries to use network properties to explain observed phenomena. This approach has been most prominent in studies that seek to find what allows ecological systems to be stable despite their complexity (May, 1974). It is surprisingly rare elsewhere in ecology.

Addressing all these types of research is impractical for a review of this length. Rather, after giving some introductory terminology, I will discuss descriptive research, explanations of network structure, and network explanations of system characteristics, with a focus on food webs and ecosystems. I will then make the case for wider use of network explanations in ecology.

2.1 Introductory Terminology

Ecosystem models or *ecosystem networks* trace the flow of matter or energy, termed the model's *currency*, through an ecosystem. A *food web* displays trophic interactions among a group of organisms. Unlike an ecosystem model, a food web usually does not include detritus and detritivores; most food webs are also less aggregated than most ecosystem models. An *interaction web* or *interaction network* can include multiple types of interactions, while a food web only includes trophic ones.

A *mutualistic network* is an interaction network depicting a relationship that benefits both parties, such as a *pollination network*, which depicts connections between plants and their pollinators, or a *common mycorrhizal network* (CMN), which depicts connections between plants and mycorrhizal fungi. The existence of common mycorrhizal networks was suggested by Newman (1988); previous work on mycorrhizae focused on pairwise interactions.

Many important terms used in network research come from graph theory. A graph consists of a set of *nodes* (or *vertices*) linked by *edges*, which may be directed or undirected. Nodes linked by an edge are termed *adjacent* and the square matrix that has a 1 in the i, j-th position if nodes i and j are adjacent and a 0 otherwise is called the graph's *adjacency matrix*. A sequence of adjacent nodes and edges in which no nodes are repeated is called a *path*; if only the first and last nodes are repeated, it is called a *cycle*; in the ecological literature (e.g. Neutel et al., 2002), the term *loop* is frequently used. (In graph theory (e.g. Agnarsson and Greenlaw, 2006), this term typically refers to an edge that starts and ends at the same node. Here, such edges will be referred to as *self-loops* and the term "loop" will be used interchangeably with "cycle", in keeping with common ecological usage.) The number of edges in a path is called its *length*. The *distance* between two nodes is the length of the shortest path between them; if no such path exists, we say that the distance is infinite. The *diameter* of a graph is the distance between the two most distant nodes in the graph.

A *component* of a graph is a set of nodes such that there is a path between any two nodes; if there is a path between all possible pairs of nodes in a graph, the graph consists of one component. A *bipartite graph* has two sets of nodes, A and B, such that nodes in A are adjacent only to those in B and vice versa. The *degree* of a node is the number of edges that link to it, with self-loops counted twice. In a directed graph, a node has an *indegree* and an *outdegree* (Agnarsson and Greenlaw, 2006).

A network is *modular* if it is made up of densely connected subnetworks that are loosely linked among themselves (Newman, 2011). Modules are also referred to as *compartments* (Pimm and Lawton, 1980, Krause et al., 2003), *communities* (Newman, 2011), or *blocks* (May, 1974).

In a *nested* network, a specialist (species with few links) interacts with species that are a subset of those with which a generalist (species with many links) interacts. In other words, in a perfectly nested network, a generalist is linked to all the species that a specialist is linked to, plus some others (Bascompte et al., 2003). It is possible to compute the degree of nestedness of a network.

In a *random graph*, vertices are connected by some random process, independently of any properties they may have. Such graphs may be contrasted with *regular graphs* or *lattices*, in which each vertex is connected to the same number of vertices. Random graphs can be traversed rapidly and have low clustering, meaning that the neighbors of linked vertices are no more likely to be linked themselves than any randomly chosen pair of vertices. On the other hand, in a regular graph, the distance between two randomly chosen points is large and clustering is high.

A *small world* network combines the high degree of clustering of a regular network with the short distances of a random network. It can be made by randomly rewiring some links in a lattice to create shortcuts between otherwise distant nodes (Watts and Strogatz, 1998) or by a preferential attachment process in which new links are more likely to involve nodes that already have many links (Barabási and Albert, 1999). In a small world network, most nodes have relatively few links but a few, termed *hubs*, are linked to large numbers of other nodes. Specifically, the distribution of degrees in a small world network follows a *power law*, an equation of the form $N = bD^{-a}$ in which D is a degree and N is the number of nodes with that degree (Csermely, 2006).

The terms introduced so far deal exclusively with network structure, but others are important when considering stocks and flows. *Boundary inputs* are inflows coming from the system's unmodeled environment; similarly, *boundary outputs* are outputs that leave the modeled system. A compartment's *throughflow* is the sum of its inflows or its outflows, which are equal at steady state. (When non-steady state methods are being used, as in Chapters 3 and 4, it becomes necessary to distinguish between the total inflow and outflow of a compartment.)

2.2 Descriptive Research

2.2.1 Structure

2.2.1.1 Food Webs

The earliest food webs were assembled by Forbes in the late nineteenth century and Pierce et al. in the early twentieth century (Forbes, 1880, Pimm et al., 1991). Early food web studies focused on describing trophic interactions in particular communities.

The first comparative studies of food web structure came in the late 1970s and revealed features of food web structure that did not vary with web size. These patterns, often referred to as *food web laws*, included scale-invariant ratios of links to species (Cohen, 1977, Briand and Cohen, 1984); fractions of top, intermediate and bottom species (Briand and Cohen, 1984); and proportions of links between different categories (top, intermediate and bottom) (Cohen and Briand, 1984). Also, omnivory and cycles were supposed to be rare and food chains were seen to be short (Pimm et al., 1991).

Food web laws faced one crucial problem: inadequate data quality. Most of the food webs used to derive them were small and all were highly aggregated (Dunne, 2006). Two papers published in 1991 made this problem clear. Polis' food web of the desert of Coachella Valley, CA had almost five times more links per species than the webs used to derive the food web laws. Food chains were longer than in these webs and looping and omnivory were common (Polis, 1991). Martinez's web of Little Rock Lake, discussed in a paper provocatively titled "Artifacts or Attributes?", was similarly detailed and also revealed looping, omnivory and long food chains. Martinez also went a step further in lumping species within the web and recovering results quite similar to those predicted by food web laws (Martinez, 1991).

2.2.1.2 Small-worldness

Since small world networks were first described (Watts and Strogatz, 1998), many ecological networks have been examined for possible small world structure. Montoya and Solé (2002) examined four food webs, including two webs of the same ecosystem at different times, and concluded that a small world model fit their degree distributions much better than a random network. However, Williams et al. (2002) looked at 16 webs and found that the degree distributions of most of the webs were not small worlds, having fewer highly connected species than would be expected in a power law distribution.

Mycorrhizal networks have also been examined for small world structure. CMNs can be conceptualized in three different ways: as plants linked by fungi (the most common way, reflecting the geometry seen in nature), as fungi linked by plants, and as bipartite networks of plants and fungi. Southworth et al. (2005) found that the mycorrhizal network of an oak savanna appeared random when modeled as trees linked by fungi but had a degree distribution consistent with a truncated power law, indicative of small-world structure, when fungal morphotypes were viewed as nodes linked by trees. It is important to note that in this study, trees were considered linked if they shared a fungal morphotype; thus, not all links in this network were present in the actual physical network underground.

2.2.1.3 Nestedness and modularity

Bascompte et al. (2003) found that plant-pollinator and seed dispersal networks were significantly nested, much more so than food webs.

Pimm and Lawton (1980) attempted to search for compartments in food webs but found none beyond those defined by habitat. However, their work was limited by the fact that data were scarce and what data they did have were low-resolution. Also, they had to use a homegrown algorithm for finding compartments. By contrast, Krause et al. (2003) were able to use a module-finding algorithm from social network research to examine five fairly large, high-resolution webs. They found compartments in three of them; in the Chesapeake Bay web, compartments corresponded to benthic and pelagic organisms. However, when reanalyzing the highly aggregated webs used in prior studies, compartments were only found in one web out of fourteen (Krause et al., 2003).

2.2.2 Flows and Dynamics

In 1942, Lindeman's article "The Trophic-Dynamic Aspect of Ecology" attempted to set out general principles of "food cycle" organization and dynamics, emphasizing energy flow (Lindeman, 1942). Production rates, efficiencies and energy flow networks came to be of great interest to ecologists, especially once radioisotope tracers became available after World War II (Hagen, 1992). Highly labor-intensive studies such as those of Odum (1957) and Teal (1962) synthesized large amounts of field and laboratory data to diagram energy flow in the ecosystems they studied. Such studies were often part of the new subdiscipline of systems ecology, which drew upon ideas and techniques from cybernetics and general systems theory. For example, Odum (1960) and Patten and Witkamp (1967), among many others, used analog computers to simulate ecosystem dynamics.

Descriptive work on flows and dynamics in ecological networks continues to this day and includes research ranging from quantitative food webs and ecosystem models (e.g. Banasek-Richter et al., 2009) to the question of whether fungal links move carbon from plant to plant in mycorrhizal networks (e.g. Robinson and Fitter, 1999, Pfeffer et al., 2004). However, the rest of this section will be devoted to one particular branch of systems ecology, environ analysis. Although this is a highly mathematical area of research, I am classifying it as largely descriptive because the role that the mathematics of environ analysis plays in the analysis of ecosystem models is analogous in many ways to that of a microscope or telescope – or H.T. Odum's "macroscope" (Odum, 1971). It is more of a tool for seeing ecosystem properties than for explaining them, although some of the properties found by environ analysis have been at least partially explained within its framework (e.g. Higashi and Patten, 1986).

2.2.2.1 Environ analysis and formalization

Environ analysis grew from the desire to formalize the concept of environment (Patten et al., 1976, Patten, 1978, Patten and Auble, 1981, Fath and Patten, 1999b). This formalization was meant to clarify the concept and link it to general systems theory in order to connect ecological and other systems. For that reason, Patten et al. (1976) drew heavily upon the state space system theory of Zadeh and Desoer (1963). To gain precision and generality, this monograph, which first outlined the ideas behind environ analysis, spent a considerable amount of time on the philosophy of causality and highly abstract representations of dynamical systems before moving on to the relatively more familiar (to ecologists) territory of networks and differential equations.

In environ analysis, an organism or ecosystem compartment is seen as the nexus of two within-system environments, termed its *input environ* and *output environ* (Patten et al., 1976, Fath and Patten, 1999b, Borrett and Freeze, 2011). The input environ of a compartment traces the flows responsible for the boundary outflow from that compartment back to the system boundary; its output environ traces the fate of boundary input to that compartment as it makes its way through the system to boundary outputs (Patten, 1978). Environs partition the flows and stocks in an ecosystem so that summing either the input environs or the output environs of all compartments reconstitutes the system's stocks and flows. An environ can and usually does include flows and compartments that are not directly linked to the focal one.

Since the original development of the environ concept, environ analysis has primarily focused on describing ecosystems. Most contemporary work in environ analysis does not deal specifically with environs. Rather, it uses related techniques and mathematical results involving power series of matrices to analyze and describe the stocks and flows in ecosystems. These techniques fall into five categories: structural analysis, flow analysis, storage analysis, utility analysis and the most recently developed one, distributed control analysis. This review will discuss the first four.

Structural analysis primarily deals with the increase of path numbers with path length. Path lengths are calculated using powers of the adjacency matrix, $\mathbf{A}$. The i, j-th entry of the matrix $\mathbf{A}^n$ gives the number of paths of length n between nodes i and j. If a graph contains cycles, the numbers of paths between nodes increases rapidly with path length; in an acyclic graph, such an increase may occur for small path lengths but must eventually stop, as the longest possible path in an acyclic graph with N nodes has length $N - 1$ (Agnarsson and Greenlaw, 2006). Since ecosystems must contain decomposers and detritus and reuse decomposed material, essentially all ecosystem models contain at least one cycle and thus have large numbers of long paths. This property is termed *network proliferation* and underlies many of the network properties and hypotheses described below.

Throughflow and storage analyses Throughflow analysis has probably been the most fruitful area of work in environ analysis. It begins with the matrix of intercompartmental currency flows, $\mathbf{F}$, oriented from columns to rows. (Self-loops, which represent storage, are not allowed in the $\mathbf{F}$ matrix.) The matrix $\bar{\mathbf{F}}$ is then defined as $\mathbf{F}$ with negative throughflows on the diagonal. Then, for a system at steady state, forward- and reverse-time ordinary differential equation descriptions of model dynamics, in matrix notation, are:

$$\frac{d\mathbf{x}}{dt} = \mathbf{0} = \bar{\mathbf{F}} \cdot \mathbf{1} + \mathbf{z} \tag{2.1}$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{0} = \bar{\mathbf{F}}^{\top} \cdot \mathbf{1} + \mathbf{y} \tag{2.2}$$

(The dynamic case will be discussed in Chapter 3.) The first equation represents time-forward dynamics generated by input $\mathbf{z}$. The second denotes reverse-time trace-back dynamics beginning at output $\mathbf{y}$,which serves as the forcing condition. In Eq. 2.2, taking the transpose of $\bar{\mathbf{F}}$ orients it backwards in time.

Throughflow analysis uses flow intensity matrices, termed $\mathbf{N}$ and $\mathbf{N}'$, that convert boundary inputs and outputs into steady-state throughflows:

$$\mathbf{T} = \mathbf{N}\mathbf{z} \tag{2.3}$$

$$\mathbf{T} = \mathbf{N}'\mathbf{y} \tag{2.4}$$

Here, $\mathbf{N} = (\mathbf{I} - \mathbf{G}_{n\times n})^{-1}$ and $\mathbf{N}' = (\mathbf{I} - \mathbf{G}'_{n\times n})^{-1}$, where $\mathbf{I}_{n\times n}$ is the identity matrix, the elements of $\mathbf{G}$ are $g_{ij} = {}^{f_{ij}}/_{T_j}$, and those of $\mathbf{G}'$ are $g'_{ij} = {}^{f_{ij}}/_{T_i}$. Both $\mathbf{G}$ and $\mathbf{G}'$ are dimensionless. Inputs, $\mathbf{z}$, outputs, $\mathbf{y}$, and throughflows, $\mathbf{T}$, all have the same dimensions, so $\mathbf{N}$ and $\mathbf{N}'$, Eqs. (2a), are dimensionless transformations from boundary flows, $\mathbf{z}$ and $\mathbf{y}$, to interior throughflows, $\mathbf{T}$. Both Eqs. 2.3 and 2.4 have infinite power series equivalents that reflect the trajectories of boundary flows over all interior pathways of all lengths traveled in reaching the points where the steady-state throughflows, $\mathbf{T}$, are registered: $\mathbf{T} = (\mathbf{I} + \mathbf{G} + \mathbf{G}^2 + \ldots + \mathbf{G}^k + \ldots)\mathbf{z}$ and $\mathbf{T} = (\mathbf{I} + \mathbf{G}' + \mathbf{G}'^2 + \ldots + \mathbf{G}'^k + \ldots)\mathbf{y}$.

Storage analysis is similar to throughflow analysis but examines the conversion of boundary inputs to stocks. System dynamics are written using the stock-normalized flow matrices $\mathbf{C}$ and $\mathbf{C}'$, in which $c_{ij} = {}^{f_{ij}}/_{x_j}$ and $c'_{ij} = {}^{f_{ij}}/_{x_i}$. (The $\mathbf{C}$ matrix is a special case of the community matrix used in population and community ecology, especially work on stability.) We then have

$$\frac{d\mathbf{x}}{dt} = \mathbf{C}\mathbf{x} + \mathbf{z} \tag{2.5}$$

$$\frac{d\mathbf{x}}{dt} = -\mathbf{C}'\mathbf{x} - \mathbf{y} \tag{2.6}$$

The $\mathbf{C}$ and $\mathbf{C}'$ matrices can also be derived from a matrix of flow probabilities, as is done in Patten (1985) and Fath and Patten (1999b). Then, at steady state, the matrices $\mathbf{S} = (-\mathbf{C})^{-1}$ and $\mathbf{S}' = (-\mathbf{C}')^{-1}$ transform inflows and outflows into stock values (Matis and Patten, 1981):

$$\mathbf{x} = \mathbf{S}\mathbf{z} \tag{2.7}$$

$$\mathbf{x} = \mathbf{S}'\mathbf{y} \tag{2.8}$$

Throughflow and storage analyses have revealed several properties of ecosystems. For example, in most ecosystem models, more currency travels over indirect pathways than direct ones (Patten, 1986, Higashi and Patten, 1989, Patten, 1990, Borrett et al., 2010). This is termed *network nonlocality* or *dominance of indirect effects*, and supports the hypothesis that control in ecosystems is also predominantly indirect. Such distributed control has become an active area of research, drawing on control theory from engineering (e.g. Schramski et al., 2007).

Network internal amplification is another property found through storage and throughflow analyses. It refers to the finding that currency introduced into one compartment will often appear more than once in that compartment (Jørgensen et al., 2007). While this has long been known to happen with elements, network environ analysis revealed that it also occurs with energy (Patten, 1985, 1986, Fath and Patten, 1999b). While such cycling appears at first glance to violate the second law of thermodynamics, no such thing happens. Rather, the fact that the net flow of energy through an ecosystem is one-way does not preclude local cycles. These cycles are like eddies in a stream. In such eddies, water can temporarily flow uphill, even though the overall flow of the stream is downhill.

Network trophic dynamics Network trophic dynamics applies the methods of environ analysis to networks of trophic interactions, with a focus on concepts, such as trophic levels and progressive efficiency, that have traditionally been important to this area of ecology. The goals of research in network trophic dynamics have been to improve or clarify concepts and describe new properties of ecosystems.

The trophic level concept has been very important to both community ecology and network trophic dynamics. The background of this idea was put into place by Semper in the late nineteenth century and Elton in the early twentieth century (Cousins, 1987); however, the trophic level concept was first explicitly articulated by Hutchinson, as quoted in Lindeman (1942). Hutchinson's trophic levels were groups of organisms, and they were discrete and sequential, although Lindeman indicated that higher trophic levels tended to blur (Lindeman, 1942).

Two revisions of the trophic level concept were put forward as solutions to the problem of blurring. One, associated with community ecology, sees an organism's trophic level as a measure of its position in a food web and resolves the blurring problem by allowing non-integer trophic levels (Levine, 1980, Williams and Martinez, 2004). The other, more common within ecosystem ecology, retains integer trophic levels and defines them as stages of energy processing (Odum, 1968). In this view, any species may occupy several trophic levels simultaneously.

Network trophic dynamics starts with the concept of discrete trophic levels as stages of energy processing and develops it using flow analysis. This *network unfolding* shows an indefinite proliferation of trophic levels (Higashi et al., 1991), which are sometimes called transfer levels, as not all flows are strictly trophic ones (Whipple and Patten, 1993, Whipple, 1998). In general, each transfer level is made up of portions of multiple compartments. The concept of progressive efficiency (the ratio of productivities of successive trophic levels), which only makes sense for discrete trophic levels, can be applied to unfolded networks (Higashi et al., 1991).

Network trophic dynamics research has described an ecosystem property that has been termed *network homogenization*. When matter or energy first enters an ecological network, it travels over pathways that are fairly well defined. However, as cycling proceeds, the amounts flowing over various paths become more equal (Higashi et al., 1993b,a, Fath and Patten, 1999a, Borrett and Salas, 2010). Other research in network trophic dynamics has characterized transfers between trophic levels using information theory (Higashi et al., 1991), partitioned currency flows into first passage and cycling (Higashi et al., 1993b), and linked environ analysis to the energy-based analyses of H.T. Odum (Patten, 1992).

Utility analysis Ecological relationships are classified by whether each species benefits from or is harmed by the relationship. For example, a competitive relationship is represented as (-,-), a mutualistic one as (+,+), and predation as (+,-). This can be extended to species that interact only indirectly. For example, in a three-species food chain, the primary producer and predator have an indirect mutualistic relationship because an increase in producer biomass benefits the predator by increasing prey abundance and and increase in predator abundance benefits the producer by reducing herbivory.

Utility analysis attempts to use flow matrices to systematically infer the ultimate relationships between species or ecosystem compartments. This is done using power series of the $\mathbf{D}$ matrix, which is defined as $d_{ij} = \frac{f_{ij} - f_{ji}}{T_i}$ and $d_{ji} = \frac{f_{ji} - f_{ij}}{T_j}$. (Since, as before, self-loops are disallowed, $d_{ii} = 0$.) Essentially, the $\mathbf{D}$ matrix describes whether a compartment obtains a net gain or loss of currency in an interaction and scales this gain or loss by the compartment's total throughflow (Patten, 1991).

The $\mathbf{D}$ matrix describes local interactions. To obtain ultimate interactions, local interactions are integrated over the network using a power series: $\mathbf{U} = \mathbf{I} + \mathbf{D} + \mathbf{D}^2 + ... = (\mathbf{I} - \mathbf{D})^{-1}$. However, $\mathbf{U}$ does not always exist, as the power series only converges when the magnitude of the dominant eigenvalue of $\mathbf{D}$ is less than one (Patten, 1991). In addition, the interpretation of the power series used to compute $\mathbf{U}$ is less straightforward than the interpretation of those used to compute $\mathbf{N}$ and $\mathbf{S}$.

Despite these difficulties, utility analysis has been used to demonstrate two related properties of ecosystems. The first, *network synergism*, is the tendency for ultimate utilities to be more positive than direct utilities (Fath and Patten, 1998), while the second, *network mutualism*, is the tendency for ultimate utility matrices to contain a greater number of positive entries than direct utility matrices (Fath, 2007). Network synergism is universal but network mutualism is not.

Utility analysis as currently practiced suffers from some conceptual difficulties. The $\mathbf{D}$ matrix is derived from the flow matrix, which is a purely descriptive, "bookkeeping" model and says nothing about the functional forms of causal relationships. Environ analysis typically assumes linear, donor-dependent dynamics, but this does not work for utility analysis.

For example, under donor-dependent dynamics, the producer and predator species in a three-species food chain cannot have a mutualistic relationship because an increase in the predator's abundance has no effect on the herbivore and, thus, no effect on the producer. Similarly, if the dynamics are purely recipient-controlled, an increase in producer biomass cannot affect the herbivore or predator. However, utility analysis demonstrates indirect mutualism in three-species food chains (Patten, 1991), implying that flow rates depend on both donor and recipient abundance. This makes good biological sense, but it is disturbing that such a substantive dynamical assumption has crept into the analysis without anyone having explicitly made it.

Another difficulty with performing utility analysis based on a purely phenomenological model is that investigating whether a species benefits from another one means asking what would happen to species i if the abundance of species j changed. The flow matrix, being purely observational, gives us no information about what would happen to the system under such an intervention or perturbation (Pearl, 2009). A model that explicitly describes the causal relationships between species or compartments is necessary.

How, then, might we perform utility analysis? A possible approach would use the community matrix, $\mathbf{K}$, which gives the coefficients of a generalized Lotka-Volterra model (the direct effects of each species on all other species) or could be extended to include other functional forms, in which case it would have to be evaluated at a stable fixed point. The community matrix gives the signs of local utility, but to perform the scaling that takes into account how important a particular flow is to a compartment, it is necessary to transpose the community matrix and multiply it by a matrix containing the state values on the diagonal:

$$\mathbf{D} = \mathbf{K}^{\top}\mathbf{x}^{*} \tag{2.9}$$

In order to find ultimate utility, it is necessary to keep in mind that utility does not propagate over the network; only influence propagates. Therefore, we must use a power series of $\mathbf{K}$ and then scale it as in Eq. 2.9 to compute utility:

$$\mathbf{U} = (\mathbf{I} + \mathbf{K} + \mathbf{K}^2 + ...)^\top \mathbf{x}^* \tag{2.10}$$

The question then becomes whether this power series is convergent. The constraints imposed upon the community matrix by an ecosystem model with a conservative currency should increase the probability of convergence. However, even if this particular approach to utility analysis fails, any subsequent attempts must obey two basic principles: utility can only be calculated for a causal model, and utility does not propagate.

2.3 Explanations of Network Structure

2.3.1 Food Webs

Two types of models have been used to explain food web structure: community and evolutionary models. In community models, species and their traits are taken as given; the model's goal is to describe the rules by which these traits result in observed network structures. In evolutionary models (e.g. Rossberg et al., 2005), species traits can change or speciation can occur. The model still needs some linking rules, but the emphasis is on the consequences of evolution. This review will only discuss community models.

The cascade model of food web structure was proposed to explain food web laws and acyclicity (Cohen and Newman, 1985). This model makes the assumption that species in a food web can be placed in order along some dimension such that a consumer can only eat species that rank lower than itself along that dimension (Cohen, 1978, Cohen et al., 1990). For reasons of stability, connectance is assumed to decline with species richness; thus, the number of links per species is constant (Pimm et al., 1991).

The niche model (Williams and Martinez, 2000) is similar to the cascade model in assigning each species a random “niche value”. However, instead of allowing each species to eat anything with a lower niche value, it assumed the each species could feed within a distribution centered on a random niche value lower than that of the consumer species (Williams and Martinez, 2000). The niche model embodies the assumption that connectance, not the number of links per species, remains constant as a food web grows in size (Martinez, 1992, 1993). It predicts a number of food web characteristics, including fractions of top, intermediate and basal species, the variances of numbers of prey and predators a species has, the mean and standard deviation of food chain length, and the prevalence of cannibalism and omnivory. Williams and Martinez’s choices of parameters to measure were not arbitrary; rather, they were related to previously identified patterns in food web structure, as described above.

The nested hierarchy model (Cattin et al., 2004), was meant to model phylogenetic constraints on species’ potential food sources, although it does not explicitly incorporate evolution. It shares a basic structure with the niche model but, unlike the cascade and niche models, can give a consumer a set of prey that are not contiguous in niche space. The nested hierarchy model gives a better fit to the data than the niche model with regard to some network features but performs worse with regard to many others (Martinez and Cushing, 2005). Other modifications of the niche model to allow non-contiguous sets of prey species have been developed, including the generalized niche model (Stouffer et al., 2006), the minimum potential niche model (Allesina et al., 2008) and the probabilistic niche model (Williams et al., 2010). The method by which the latter two models were tested against data differs in important ways from how earlier models were tested and will be discussed below.

2.3.2 Mutualistic Networks

The structure of mutualistic networks has been studied using models similar to those used for food webs, albeit less work has been done in this area. Pires et al. (2011) actually applied the niche and cascade models, modified for bipartite networks, to plant-pollinator and plant-frugivore networks and obtained a good fit, which they explained with the fact that these mutualisms are based on consumer-resource relationships (Pires et al., 2011). An earlier paper, Santamaria and Rodriguez-Gironés (2007), tested models incorporating trait complementarity, barriers, or both. They concluded that both were involved in structuring plant-pollinator networks; however, they tested the models in an unusual way. Rather than comparing individual model realizations to data, Santamaria and Rodriguez-Gironés regressed network nestedness against connectance and number of links using real and simulated data and judged the goodness of fit of a model by how closely it approximated this regression.

2.3.3 How Should We Evaluate Models of Network Structure?

No model describes all aspects of the entity it represents. If it did, it would be as complicated as nature itself and therefore useless as an aid to understanding. This simple realization has important implications for how models should be tested against data.

Articles such as Williams and Martinez (2000) compare models to data by computing a number of summary statistics and comparing modeled values to observed ones. However, recent papers such as Allesina et al. (2008) and Williams et al. (2010) use a likelihood-based approach that evaluates models based on how well they reproduce the presence or absence of individual links in a web. Similarly, in “Current food web models cannot explain the overall topological structure of observed food webs”, Fox (2006) quantified a measure called structural stability for real, niche model, and cascade model food webs in order to test the performance of the models in predicting the “details” of web structure. Fox starts out by saying that food web models attempt to explain observed patterns in web structure, which is correct, but quickly slides to “the rules governing food web topology” (Fox, 2006), apparently without any awareness of the huge step he has just taken. While these papers share the seemingly unobjectionable goal of subjecting models to more rigorous tests, they may focus too much on the details of food web structure.

The structure of real food webs is determined by both general factors such as body size and nutrient stoichiometry and a myriad of natural history details ranging from the secondary metabolite composition of plants to the times of day when particular animals are active. This makes the food web itself a poor object of explanation, as a small change in the biology of the species in the web will change the detailed structure of the network. To understand why this makes a given food web a poor object of explanation, consider a speeding driver who crashes into another vehicle. When blamed for having caused the accident by speeding, he replies that if he hadn't eaten breakfast that morning, he also would not have had that accident, so there is no reason to single out his speeding as having caused the crash. The driver is defining "that accident" very narrowly so that a small change would result in him not having had "that accident", even if he would likely have had a similar one (Garfinkel, 1981).

Just as we are not interested in why the speeding driver's accident happened at the exact place and time that it did, we do not really want to know why a particular food web has the precise structure that it does.[1] Such an explanation, if it existed, would be too detailed to be comprehensible and would tell us nothing about food webs other than the one under study. A scientifically useful model should explain the *regularities* of food web structure, not the accidental features of particular webs. Therefore, such models are better judged by their prediction of ecologically interesting summary statistics or network descriptors, as in the older literature, than by their prediction of the presence or absence of individual links, as in the likelihood approach. Structural stability may well be a network

---

[1] The biochemist and writer Nick Lane makes the same point when he writes, "The quest for the origin of life is not an attempt to reconstruct what happened at 6:30 a.m. on Thursday morning in the year 3,851 million BC, but for the general rules that must govern the emergence of any life, anywhere in the universe, and especially on our planet, the only example we know" (Lane, 2009).

characteristic worth predicting, but we should test how well food web models predict this or other characteristics only if we are interested in the characteristics themselves rather than using arbitrary characteristics as a way of quantifying the details of web structure.

2.4 Network Explanations

In the 1960s, two working-class Boston neighborhoods were the subjects of urban renewal plans. The residents of Charlestown were able to successfully organize to resist the plans; those of the West End were not. Comparing sociological studies of the two neighborhoods (as well as others), Granovetter (1973) hypothesized that West End residents were unable to organize because their community was fragmented into tight cliques with few bridges, while the Charlestown residents, who had more opportunities for forming ties between people who were not relatives or close friends, could effectively organize against the city's plans.

Granovetter's explanation of the different responses of the two neighborhoods is of a particular kind. It uses an aspect of *network structure* (the abundance or scarcity of weak ties) to explain an observed *macro-level property* (the ability or lack of ability to organize). In another example, Eagle et al. (2010) found a strong positive correlation between the diversity (but not volume) of phone contacts people in a community had and the prosperity of that community. They interpreted this finding to indicate that diversity of social contacts helps community development, helping to explain it.

2.4.1 Diversity and Stability

In ecology, network explanations have been most prominent in research on the relationship between diversity and stability. A good history of this line of work is given by Justus (2008), so I will only address its network-related aspects.

The idea that species diversity raises community stability was first articulated in the late

1940s and 1950s. MacArthur (1955) defined stability as the tendency of a perturbation in the numbers of one species in a food wed to diminish as it was propagated to other species in the web. His conception of stability was quantitative – an ecosystem could be more or less stable, not just stable or unstable – and based on energy flow considerations. MacArthur reasoned that if a food web had many links, unusual concentrations of energy in one compartment would be rapidly distributed throughout the web. Furthermore, if a species had many prey items, an increase in its abundance would affect each prey species only a little. Thus, species richness would stabilize an ecosystem.

Despite some empirical and theoretical work suggesting that the relationship between ecosystem complexity and stability was not a simple one, this consensus lasted through the early 1970s. In 1972, however, Robert May published a paper in *Nature* entitled, "Will a large complex system be stable?" (May, 1972). This was followed by *Stability and Complexity in Model Ecosystems* (May, 1974). Both the book and the paper showed that, in classic Lotka-Volterra community models with randomly chosen interaction coefficients, stability actually became less likely as species were added to the system. Increasing connectivity (the fraction of possible connections among community members that actually exist) or mean interaction strength also lowered the probability that a system would be stable.

At this point, it is worth comparing May's and MacArthur's work on complexity and stability. MacArthur's argument was qualitative – only one equation appears in the paper, and that is for a purported index of stability – but his conception of stability was a quantitative, graduated one, albeit vaguely defined. MacArthur reasoned in terms of energy and the dissipation of unusually high amounts of energy in a compartment. May, on the other hand, used the more abstract interaction coefficients of Lotka-Volterra models and defined stability as *local* or *Lyapunov stability*. Under this definition, a system is stable if, after an infinitesimal perturbation, all populations return to equilibrium, defined in the mathematical sense of no change in population numbers. If the system is not at equilibrium to begin with, this stability concept does not apply. However, a field ecologist would be extremely surprised to see all populations remain absolutely constant year after year, even in the absence of disturbance. Living systems change, although they do so within bounds. Most of the post-May attempts to show how complex communities can be stable have focused on one or another of his randomness assumptions; however, the simple equating of stability with absence of change is equally open to question.

Despite his repeated use of the word "population", MacArthur's conception of food webs was an ecosystem one, couched in terms of energy flow rather than the more abstract interaction coefficients of Lotka-Volterra models. MacArthur assumed that there was a limit to how much energy a predator population of a given size could use, so there could not be a large number of strong links. May's Lotka-Volterra models were subject to no such constraint.

Other ecologists have studied the effects of interaction parameters to determine how real-life complex ecosystems can persist. Yodzis showed the importance of parameter choice in forty real food webs, assigning interaction coefficients in a way guided by the real biology of each species (Yodzis, 1981). Yodzis performed a test in which all interaction parameters were randomly permuted while preserving topology and self-limitation. Almost none of the resulting models were stable. Switching pairs of coefficients rather than individual numbers produced a similar but somewhat weaker result. Clearly, real ecological interaction strengths are not random and this nonrandomness is biologically significant.

Interaction strengths again come to the forefront in "Weak trophic interactions and the balance of nature" (McCann et al., 1998). The models discussed in this paper, while similar to Lotka-Volterra models in that population interactions are controlled by the product of predator and prey abundances, incorporate logistic growth of the basal species and consumer preference for one prey type over another, which results in weak links. These weak links stabilize trophic interactions in a way similar to MacArthur's original conception. The introduction of weak links changes chaotic dynamics to oscillatory ones and oscillations to equilibria.

McCann et al. do not address any form of stability; however, it has been found that selectively removing weak links from an ecosystem model increased the time it took to approach equilibrium after a disturbance (Pinnegar et al., 2005). Interestingly, Allesina and Tang (2012) analytically found that, in Lotka-Volterra predator-prey models, weak links should actually be destabilizing, but their use of models without a saturating functional response and the fact that they assigned prey and predator interaction strengths indepently of each other limits the biological applicability of their work. Their focus on local stability is also problematic for the reasons described above.

2.5 Conclusions and Future Directions

One of the primary goals of science is the explanation of natural phenomena. Some of the most important theories in all of science gained their status because of the explanations they provided. Newton's theory of gravity explained Kepler's laws of planetary motion and gained acceptance because of its explanatory successes (Toulmin, 1961, Putnam, 1991). The theory of evolution by natural selection explained why species change, as well as other facts such as biogeographical distributions and patterns of embryological development. The theory of plate tectonics explained correspondences between the shapes of continents, as well as the locations of earthquakes and volcanoes and the distributions of fossils.

In the social sciences, network structure has been used to explain various phenomena, as discussed above. In a more ecologically relevant example, Rothenberg et al. (1998) describe a population at high risk for contracting HIV (prostitutes, injecting drug users, and their sexual partners) in which HIV transmission was, in fact, low. This is explained by the fact that the social and sexual networks in this population were highly fragmented, with no highly connected subgroups (Rothenberg et al., 1998). By contrast, a syphilis outbreak in another population was associated with an increase in network cohesiveness (Potterat et al., 1999).

Network-oriented ecologists have come up with a plethora of network properties and measures, including degree distributions, mean trophic level, Finn's cycling index, ascendency, the ratio of indirect to direct flows, network homogenization, network mutualism, and many others (Borrett and Salas, 2010, Fath et al., 2001, Ulanowicz, 1997). The drivers of these properties have been studied, but the properties themselves have, with the exception of research on stability, not been used as explanations themselves. If such uses are possible, as I believe they are, this is a tremendous missed opportunity; if they are not possible, empirically-oriented ecologists may justifiably ask why they should care about all this network stuff.

What kinds of community and ecosystem properties might network characteristics be able to explain? Some possibilities include invasibility, susceptibility to eutrophication and biomagnification (a high amount of cycling should increase biomagnification), and differences in ecosystems' resilience and resistance to perturbations. For example, it would be very interesting to find out whether communities that withstand extinctions well have higher proportions of indirect energy flow than less robust systems. Such work will require close collaboration between empirical and theoretical ecologists; ideally, we should try to do both kinds of research.

I am not arguing that explanation is the only thing network ecologists should be doing. Descriptive work continues to be important, as does theoretical modeling. Yet, we are in the business of explaining nature and network explanations may be the most important thing systems ecologists have to offer to the rest of the discipline.

Chapter 3

Dynamic Environ Analysis of Compartmental Systems: A Computational Approach[1]

---

[1] Shevtsov, J., Kazanci, C. and B.C. Patten. 2009. *Ecological Modelling.* 220:3219-3224. Reprinted here with permission of the publisher.

Abstract

Ecosystems are often modeled as stocks of matter or energy connected by flows. Network environ analysis (NEA) is a set of mathematical methods for using powers of matrices to trace energy and material flows through such models. NEA has revealed several interesting properties of flow–storage networks, including dominance of indirect effects and the tendency for networks to create mutually positive interactions between species. However, the applicability of NEA is greatly limited by the fact that it can only be applied to models at constant steady states. In this paper, we present a new, computationally oriented approach to environ analysis called dynamic environ approximation (DEA). As a test of DEA, we use it to compute compartment throughflow in two implementations of a model of energy flow through an oyster reef ecosystem. We use a newly derived equation to compute model throughflow and compare its output to that of DEA. We find that DEA approximates the exact results given by this equation quite closely – in this particular case, with a mean Euclidean error ranging between 0.0008 and 0.21 – which gives a sense of how closely it reproduces other NEA-related quantities that cannot be exactly computed and discuss how to reduce this error. An application to calculating indirect flows in ecosystems is also discussed and dominance of indirect effects in a nonlinear model is demonstrated.

3.1 Introduction

Compartment models (Matis et al., 1979) are widely used to represent ecological networks of *stocks*, x_i $(i = 1, 2, .., n)$, and *flows*, f_{ij} $(i, j = 1, 2, .., n)$, of conserved substances (energy or matter). The flows are generated by boundary *inputs*, z_j, and they terminate in boundary *outputs*, y_i . *Throughflows* are the sums of inflows, T_i^{in}, and outflows, T_i^{out}, to and from each stock. Within-model environments of the compartments are *environs* (Patten, 1978). These may be found using the system's mathematical description by *network environ analysis*

(NEA), a set of methods derived from Leontief (1936, 1966) input–output analysis. NEA has revealed several interesting properties of flow–storage networks, including dominance of indirect effects (Patten, 1984, Higashi and Patten, 1989) and the tendency for networks to create mutually positive interactions between species (Patten, 1991).

At least three aspects of dynamical behavior limit the applicability of present NEA methods. (1) The methodology can only be applied to models at constant steady states where inputs balance outputs. This greatly limits the range of applicability because (2) not all models reach constant steady states, and (3) those that do may also have significant, but unanalyzable, transient behavior. Previous attempts to respond to these limitations and develop methods for non-steady-state linear (Hippe, 1983) as well as nonlinear (Hallam and Antonios, 1985) systems have not found use, in part because of their mathematical difficulty.

This paper describes a computational approach to dynamic environ analysis. Like NEA, the dynamic methodology can be applied to any compartment model that satisfies two properties. First, either all compartments that have an input must have a boundary output or, failing that, every block of compartments that receives an input must have a boundary output. Second, at least one compartment must receive input from outside the system to prevent system descent to the zero state (although zero-input transient dynamics from a nonzero initial state may be of interest, and could be analyzed using DEA).

3.2 The method

3.2.1 Overview of standard environ analysis

For a compartmental system, let $\mathbf{x}_{n\times 1} = (x_i)$, $\mathbf{z}_{n\times 1} = (z_j)$, and $\mathbf{y}_{n\times 1} = (y_i)$ be stock, input, and output vectors, respectively; let $\mathbf{1}_{n\times 1}$ be a vector of ones, and $\mathbf{F}^\top$ the transpose of the matrix of flows, $\mathbf{F}_{n\times n} = (f_{ij})$. We define $\bar{\mathbf{F}}$ as the flow matrix $\mathbf{F}$ with negative throughflows on the diagonal, so $\bar{f} = f_{ij}$ for $i \neq j$ and $\bar{f} = -T_i$. Then, for a system at steady state, input- and output-driven ordinary differential equation descriptions of model dynamics, in matrix notation, are

$$\frac{d\mathbf{x}}{dt} = 0 = \bar{\mathbf{F}}\cdot\mathbf{1} + \mathbf{z} \tag{3.1a}$$

$$0 = -\bar{\mathbf{F}}^{\top} \cdot \mathbf{1} - \mathbf{y} \tag{3.1b}$$

The first equation represents time-forward dynamics generated by input, $\mathbf{z}$. The second denotes reverse-time trace-back dynamics beginning at output, $\mathbf{y}$, which serves as the forcing condition. (The flows $\mathbf{z}$ and $\mathbf{y}$ may be termed boundary flows.) In Eq. 3.1b, taking the transpose of $\bar{\mathbf{F}}$ orients it to backward movement of time, signified by the negative signs of both terms.

Standard NEA converts boundary inputs (in output-environ analysis) and outputs (in input-environ analysis) into steady-state throughflows, $\mathbf{T}_{n\times 1} = (T_i^{in}) = (T_i^{out})$, and storages (stocks), $\mathbf{x}_{n\times 1} = (x_i)$, employing flow intensity matrices, $\mathbf{N}_{n\times n}$ and $\mathbf{N}'_{n\times n}$ for throughflow analysis, and $\mathbf{S}_{n\times n}$ and $\mathbf{S}'_{n\times n}$ for storage analysis:

$$\mathbf{T} = \mathbf{N}\mathbf{z} = \mathbf{N}'\mathbf{y} \tag{3.2a}$$

$$\mathbf{x} = \mathbf{S}\mathbf{z} = \mathbf{S}'\mathbf{y} \tag{3.2b}$$

Here, $\mathbf{N} = (\mathbf{I}{-}\mathbf{G}_{n\times n})^{-1}$, $\mathbf{N}' = (\mathbf{I}{-}\mathbf{G}'_{n\times n})^{-1}$, $\mathbf{S} = -\mathbf{C}^{-1}_{n\times n}$, and $\mathbf{S}' = -\mathbf{C}'^{-1}_{n\times n}$, where $\mathbf{I}_{n\times n}$ is the multiplicative identity matrix, $g_{ij} = f_{ij}/T_j$ and $c_{ij} = f_{ij}/x_j$. Both $\mathbf{G}$ and $\mathbf{G}'$ are dimensionless, while $\mathbf{C}$ and $\mathbf{C}'$ have the dimensions of reciprocal time; note that $\mathbf{C}$ is the familiar "community matrix" used in population and community ecology.

Inputs, $\mathbf{z}$, outputs, $\mathbf{y}$, and throughflows, $\mathbf{T}$, have the same dimensions, therefore $\mathbf{N}$ and

$\mathbf{N}'$, Eq. 3.2a, are dimensionless transformations from boundary flows, $\mathbf{z}$ and $\mathbf{y}$, to interior throughflows, $\mathbf{T}$. Both Eqs. (2a) and (2b) have infinite power series equivalents reflect trajectories of the boundary flows over all interior pathways of all lengths traveled in reaching the points where the steady-state throughflows, $\mathbf{T}$, are registered. For Eq. 3.2a, these series are

$$\mathbf{T} = [\mathbf{I} + \mathbf{G} + \mathbf{G}^2 + \cdots + \mathbf{G}^k + \cdots]\mathbf{z} \tag{3.3a}$$

$$= [\mathbf{I} + \mathbf{G}' + \mathbf{G}'^2 + \cdots + \mathbf{G}'^k + \cdots]\mathbf{y} \tag{3.3b}$$

3.2.2 The dynamic case

The equation that governs the dynamics of a single compartment k is

$$\frac{dx_k}{dt} = T_k^{in}(t) - T_k^{out}(t) \tag{3.4}$$

where $T^{in}(t)$ and $T^{out}(t)$ are functions that represent rates of input to and output from compartment k at time t. Note that T in (t) is a k combination of environmental and inter-compartmental flows:

$$T_k^{in} = \sum_{i=1}^{n} f_{ki}(t) + z_k(t) \tag{3.5}$$

Combining Eqs. 3.4 and 3.5, we get, for $i \neq k$,

$$\sum_{i=1}^{n} f_{ki}(t) + z_k(t) = T_k^{out} + \frac{dx_k}{dt} \tag{3.6}$$

As before, we define $\mathbf{G}$, the flow matrix normalized with respect to throughflows (T_k^{out}), as $g_{ik} = f_{ik}/T_{out}$. Replacing $\mathbf{F}$ with $\mathbf{G}$ in Eq. 3.6, we get

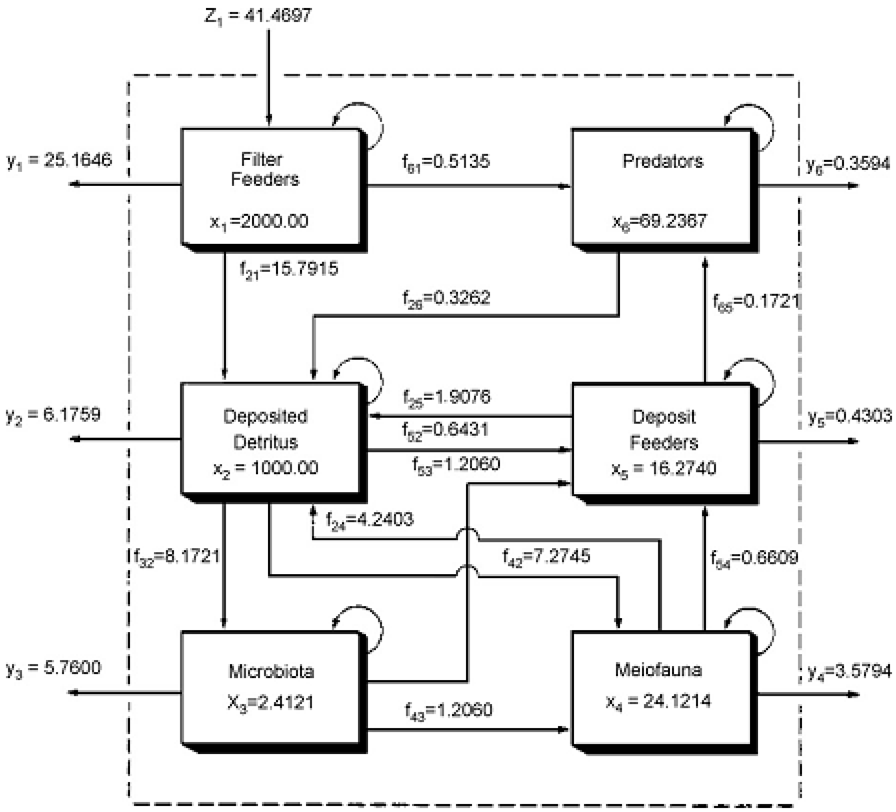


Figure 3.1: Energy flows in an oyster reef ecosystem. The stock and flow values are for a constant steady state. Note that only Compartment 1 receives direct boundary inflow. From Patten (1986).

$$z_k - \frac{dx_k}{dt} = T_k^{out} - \sum_{i=1}^{n} g_{ki} T_i^{out} \tag{3.7}$$

Using matrix notation, the equation above can be expressed as follows:

$$\mathbf{z} - \frac{d\mathbf{x}}{dt} = (\mathbf{I} - \mathbf{G})\mathbf{T}^{out} \tag{3.8}$$

Step	Description	Implementation in oyster model analysis
1	Numerically simulate the system using discrete time steps, δt, to obtain stocks, $x(t_m)$	$\delta t = 0.025$ day
2	At time steps $\Delta t \geq \delta t$, compute inputs $\mathbf{z}(t_m)$, outputs $\mathbf{y}(t_m)$, interior flows $\mathbf{F}(t_m)$ and throughflows $\mathbf{T}(t_m)$	$\Delta t = 0.25$ day (one time unit in the empirical model)
3	Letting $\mathbf{V}_{NEA}$ denote any of Eq. 3.2 transformation matrices ($\mathbf{N}$, $\mathbf{N}'$, $\mathbf{S}$, or $\mathbf{S}*$) relating these quantities, derive corresponding DEA versions, $\mathbf{V}_{DEA}$	The $\mathbf{N}_{DEA}$ matrix was calculated using a moving window of length 20
4	Perform any of the customary NEA computations using $\mathbf{V}_{DEA}$ instead of $\mathbf{V}_{NEA}$ matrices	As a test of the method, T was calculated using the equation $\mathbf{T} = \mathbf{N}_{DEA}\mathbf{z}$. Indirect flows were calculated as $\mathbf{N}_{DEA} - \mathbf{I} - \mathbf{G}$.

Table 3.1: Steps of dynamic environ approximation and illustration using oyster reef model.

Assuming that the matrix $\mathbf{I}$–$\mathbf{G}$ is invertible and $\mathbf{N} = (\mathbf{I}–\mathbf{G})^{-1}$ as before, we get

$$\mathbf{N}(\mathbf{z} - \frac{d\mathbf{x}}{dt}) = \mathbf{T}^{out} \tag{3.9}$$

Note that at steady state, $d\mathbf{x}/dt = 0$ and the above equation reduces to Eq. 3.2a.

We will now show how $\mathbf{N}$ is obtained in the dynamic case, where $d\mathbf{x}/dt \neq 0$.

3.2.3 Dynamic environ approximation

The constant matrices $\mathbf{G}$ and $\mathbf{G}'$ of static NEA do not reflect the reality that systems and their flow coefficients change over time, including the (infinite) time required for the power series in Eq. 3.3a to become equal to the transformation matrices $(\mathbf{I}–\mathbf{G})^{-1}$ and $(\mathbf{I}–\mathbf{G}')^{-1}$ of

Eqs. 3.2. In NEA, the matrix powers in Eq. 3.3a are interpreted as corresponding to pathway lengths, implying that (as $k \to \infty$) all pathways of all lengths are utilized in the limit in bringing $\mathbf{T}$ (Eqs. 3.3) to its measured or modeled value. In throughflow analysis, *paths* are pathways lacking self-loop subsequences, $...i \to i \to ... \to i...$. If each adjacent link, denoting a pathway of length 1, is associated with a discrete time of passage, $\Delta t = t_{m+1} - t_m$, then the time required to traverse pathways of lengths $m = 0, 1, 2, ..., \infty$ is $m\Delta t$. Therefore, matrix powers may also be viewed as representing numbers of time steps (Patten, 1985, Patten et al., 1990). This interpretation is helpful in understanding dynamic environ approximation. DEA involves four computational steps:

Step 1. Generate a numerical solution of the system differential equations (Eq. 3.1a or 3.1b) using discrete computational time steps, δt, to obtain stocks, $\mathbf{x}(t_m)$. This computational interval may be constant or time-varying; in the latter case, it will be necessary to interpolate $\mathbf{x}(t_m)$ for all integer values of m until the end of the time series.

Step 2. At sampling times $\Delta t = n\delta t$, where n is an integer, compute from the simulated values at times t_m the NEA quantities indicated in Eqs. 3.1a-3.2b: inputs $\mathbf{z}(t_m)$, outputs $\mathbf{y}(t_m)$, interior flows $\mathbf{F}(t_m)$ and throughflows $\mathbf{T}(t_m)$.

Step 3. Letting $\mathbf{V}_{NEA}$ denote any of the Eq. 3.2 transformation matrices ($\mathbf{N}$, $\mathbf{N}'$, $\mathbf{S}$, or $\mathbf{S}'$) relating these quantities, derive corresponding DEA versions, $\mathbf{V}_{DEA}$.

Step 4. Perform any of the customary NEA computations using $\mathbf{V}_{DEA}$ instead of $\mathbf{V}_{NEA}$ matrices.

Integer powers, m, of any scalar or matrix quantity, say $\mathbf{W}$, correspond to $m-1$ repeated multiplications of that quantity: $\mathbf{W}^m = \mathbf{W} \cdot \mathbf{W} \cdot ... \cdot \mathbf{W}$ (m terms). The innovation behind DEA is the recognition that this makes it possible to substitute a non-constant, time- and pathway-varying product series for each constant-generated time- and pathway-varying term of the NEA power series. Thus, if the generalized NEA form of the power series in Eq. 3.3a is

$$\mathbf{V}_{NEA} = \mathbf{W}^0 + \mathbf{W}^1 + \mathbf{W}^2 + \mathbf{W}^3 + \cdots + \mathbf{W}^m + \ldots \tag{3.10a}$$

$$= \mathbf{I} + \mathbf{W} + \mathbf{W}\cdot\mathbf{W} + \mathbf{W}\cdot\mathbf{W}\cdot\mathbf{W} + \cdots + [\mathbf{W}\cdot\mathbf{W}\cdot\ldots(m\text{ terms})\ldots\cdot\mathbf{W}] + \ldots \tag{3.10b}$$

then the corresponding DEA form can be written as:

$$\mathbf{V}_{DEA}(t_0) = \mathbf{V}_0 + \mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3 + \cdots + \mathbf{V}_m + \ldots \tag{3.11a}$$

This expression can be expanded to get:

$$\mathbf{V}_{DEA}(t_0) = \mathbf{W}_0 + [\mathbf{W}_0 \cdot \mathbf{W}_1] + [\mathbf{W}_0 \cdot \mathbf{W}_1 \cdot \mathbf{W}_2] + \cdots + [\mathbf{W}_0 \cdot \mathbf{W}_1 \cdot \mathbf{W}_2 \cdot \ldots \cdot \mathbf{W}_m] + \ldots \tag{3.11b}$$

where $\mathbf{W}_0 = \mathbf{I}$. Truncation after $m + 1$ terms gives the approximation:

$$\mathbf{V}_{DEA}(t_0) \approx \mathbf{W}_0 + [\mathbf{W}_0\cdot\mathbf{W}_1] + [\mathbf{W}_0\cdot\mathbf{W}_1\cdot\mathbf{W}_2] + \cdots + [\mathbf{W}_0\cdot\mathbf{W}_1\cdot\mathbf{W}_2\cdot\ldots\cdot\mathbf{W}_m] \tag{3.11c}$$

where $\mathbf{V}$ and $\mathbf{W}$ represent any appropriately related pair of NEA matrices. Here, $m + 1$ is the length of the longest product in the series as well as the number of terms in the sum. In general,

$$\begin{aligned}\mathbf{V}_{DEA}(t_k) =& \mathbf{W}(t_k) + [\mathbf{W}(t_k)\cdot\mathbf{W}(t_k + 1)] + [\mathbf{W}(t_k) \cdot \mathbf{W}(t_{k+1}) \cdot \mathbf{W}(t_{k+2})] + \cdots \\ &+ [\mathbf{W}(t_k) \cdot \mathbf{W}(t_{k+1}) \cdot \ldots \cdot \mathbf{W}(t_{k+m})]\end{aligned} \tag{3.11d}$$

These calculations can be thought of as stepping a moving window of fixed length m along the simulated dynamics at a fixed sampling interval, Δt, which must be at least as large as the numerical integration step size. At each sampling step, a value for $\mathbf{V}_{DEA}(t_k)$ is calculated. The interpretation of $\mathbf{V}_{DEA}$ is simplest when $\Delta t = 1$.

The forms for $\mathbf{V}_{DEA} = \mathbf{N}$ and $\mathbf{V}_{DEA} = \mathbf{N}'$ (with $\mathbf{W}_0 = \mathbf{G}_0 = \mathbf{G}'_0 = \mathbf{I}$ omitted from the $m > 1$ terms, as multiplication by $\mathbf{I}$ does not affect the results) are as follows:

$$\mathbf{N}_{DEA}(t_k) = \mathbf{I} + \mathbf{G}(t_1) + \mathbf{G}(t_1)\cdot\mathbf{G}(t_2) + \mathbf{G}(t_1)\cdot G(t_2)\cdot G(t_3) + ... \tag{3.12a}$$

$$\mathbf{N}'_{DEA}(t_k) = \mathbf{I} + \mathbf{G}(t_1) + \mathbf{G}(t_1)\cdot\mathbf{G}(t_2) + \mathbf{G}(t_1)\cdot\mathbf{G}(t_2)\cdot\mathbf{G}(t_3) + ... \tag{3.12b}$$

where $\mathbf{G}$ is the matrix of flows normalized by donor throughflows and $\mathbf{G}'$ is the matrix of flows normalized by recipient throughflows. Using this methodology, non-steady- or steady-state analyses can be performed, and a dynamic analysis applicable to nonlinear as well as linear systems becomes possible. Standard NEA becomes a special case of the more general DEA approach, which, in principle, becomes arbitrarily exact as $t \to \delta t$ and $\delta t \to 0$. That is, the approximations of Eq. 3.11 can be improved by re-simulating the dynamical model with coefficients re-calculated using a smaller time unit (e.g. hours instead of days). This is particularly important for systems that exhibit high-frequency dynamics.

There is a trade-off involved in increasing window size (m). Notice that $\mathbf{V}(t)$ is defined at a particular time, t, but computed using $\mathbf{W}(t+1)$, $\mathbf{W}(t+2)$ and other future values. This creates an unavoidable error, and introducing more distant time points will increase this error, while decreasing the error due to truncation of the infinite series defining $\mathbf{V}(t)$. Note that terms added to the end of Eq. 3.11d will be very small. In general, there is little gain from using a window size larger than about 20.

3.2.4 Numerical test of DEA methodology

Dame and Patten (1981) modeled energy flow in an intertidal oyster reef in South Carolina, USA (Fig. 3.1). This model has one nonzero boundary input, z_1, six compartments (x_1 to x_6) each dissipating energy to nonzero outputs (y_1 to y_6), and twelve empirically measured internal flows (f_{ij}, $i, j = 1, 2, ..., 6$). The flow units are kcal/m^2/day and the stocks kcal/m^2. Two implementations of this model, described below, were used to test dynamic environ approximation methods. Table 3.1 gives a summary of the analysis.

A linear time-forward model (Eq. 3.1a) was formulated by defining interior flows as scalar multiples of the donor compartments, $f_{ij} = c_{ij}x_j$. In Step 1 the model was simulated using EcoNet (Kazanci, 2007). A 200-time unit (50 day) run with the initial filter feeder stock, $x_1(t_0)$, displaced from 2000 to 4000 kcal/m^2 is shown in Fig. 3.2a. In Step 2, stock vectors, $\mathbf{x}(t)$, and community matrices, $\mathbf{C}(t)$, were assembled for each sample time step of $\delta t = 0.1$ day. Using the $\mathbf{C}$ matrices, $\mathbf{F}(t)$ (Eq. 3.1a), and $\mathbf{G}(t)$ matrices were computed for each sampling time in Step 3, then used to compute $\mathbf{N}_{DEA}$ (Eq. 3.12a) with a window size of 20.

A nonlinear, mass-action version of the Fig. 3.1 model was formulated by making flows functions of the product of the donor and recipient stocks, $f_{ij} = a_{ij}x_ix_j$. Flows to Detritus (x_2) remained donor-dependent, as in the linear model. The same computations as described for the linear model above were performed. For both models, the $\mathbf{N}_{DEA}$ matrix was used to calculate the throughflow vector, $\mathbf{T}$ (Eq. 3.9). Throughflow for each compartment was also computed directly as the sum of outflows in the dynamic simulation program Berkeley Madonna 8.0.1 (www.berkeleymadonna.com). Since the analysis presented here is forward-looking (i.e., Eq. 3.12a), $\mathbf{T}i$ was defined as the sum of outflows from compartment i; a backward-looking analysis (i.e., Eq. 3.12b) would have used the sum of inflows to compartment i. Error was calculated using the Euclidean norm:

$$Error = \sqrt{(T_1^{calc} - T_1^2)^2 + ... + (T_n^{calc} - T_n^{actual})^2} \tag{3.13}$$

(Here, n is the number of compartments in the model.) The error thus defined was cal-

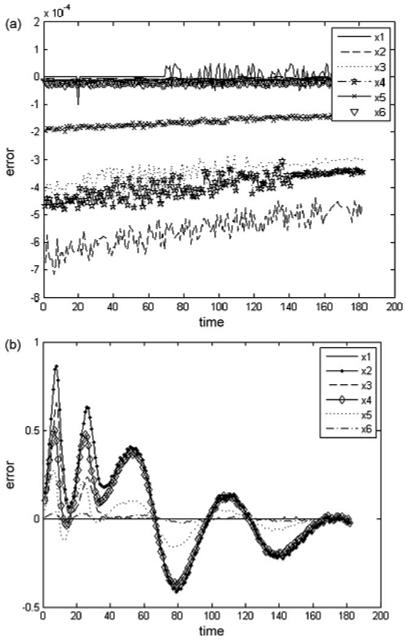


Figure 3.2: Difference, in kcal/day/m^2 , between actual and calculated throughflows in linear (a) and nonlinear (b) simulations of the oyster reef model. Error is the Euclidean distance between a throughflow vector computed with the $\mathbf{N}_{DEA}$ matrix (Eq. 3.9) and one computed as the sum of outflows in the dynamic simulation. Compartment names $(x_1, ..., x_6)$ are explained in Fig. 3.1.

culated for ten randomly selected time points in both model implementations. The mean error in the linear model was 0.0008 with a standard deviation of 0.0004; that of the nonlinear model was 0.2171 with a standard deviation of 0.11. Differences between actual and calculated throughflows are displayed in Fig. 3.2.

3.3 Application to indirect effects

Output from the linear and nonlinear dynamic oyster reef model implementations was analyzed to compute the fraction of flow between pairs of compartments that traveled over pathways of lengths greater than one (indirect flow fraction). This quantity was calculated by dividing entries in $\mathbf{N}_{DEA} - \mathbf{I} - \mathbf{G}$, which isolates indirect flows, by the corresponding entries in $\mathbf{N}$, which represents total flows. If there is no directed path of any length between two compartments, the ratio is undefined; in this case, it was arbitrarily set equal to zero. The results of this calculation for the linear model are shown in Fig. 3.3b; those for the nonlinear model appear in Fig. 3.4b–e. The Matlab function used to do these calculations is given in Appendix A.

In the linear model, direct to indirect flow ratios remained constant as the system evolved; in the nonlinear model, they varied with time. The mean and median values for the two models were similar: between 0.45 and 0.6. Typically, direct to indirect flow ratios in the nonlinear model changed gradually and, over the 45 days simulated, underwent proportionately much less change than stock values (Fig. 3.4b–e). This relationship should be explored in future research, as should the constancy of indirect flow fractions in linear models.

3.4 Discussion

The dynamic environ approximation approach described in this paper potentially has a broad range of applications. Here, we have described the approach and given an example of its accuracy.

Previous attempts to develop a dynamic environ analysis were primarily analytical (Hippe, 1983, Hallam and Antonios, 1985). (Hippe's approach appears related to the dynamic inverse in input–output analysis (Leontief, 1970, Kendrick, 1972, Johnson, 1985, ten Raa, 2006) and deeper exploration of the relationship between the two methodologies may prove worthwhile.) The strength of DEA lies in the fact that, like NEA, it makes no assumptions about the underlying dynamics of the model being analyzed. Although two simulation models were used in the present instance to produce the background data for DEA, the analysis could also proceed based on purely empirical time series data gathered in context of a defined

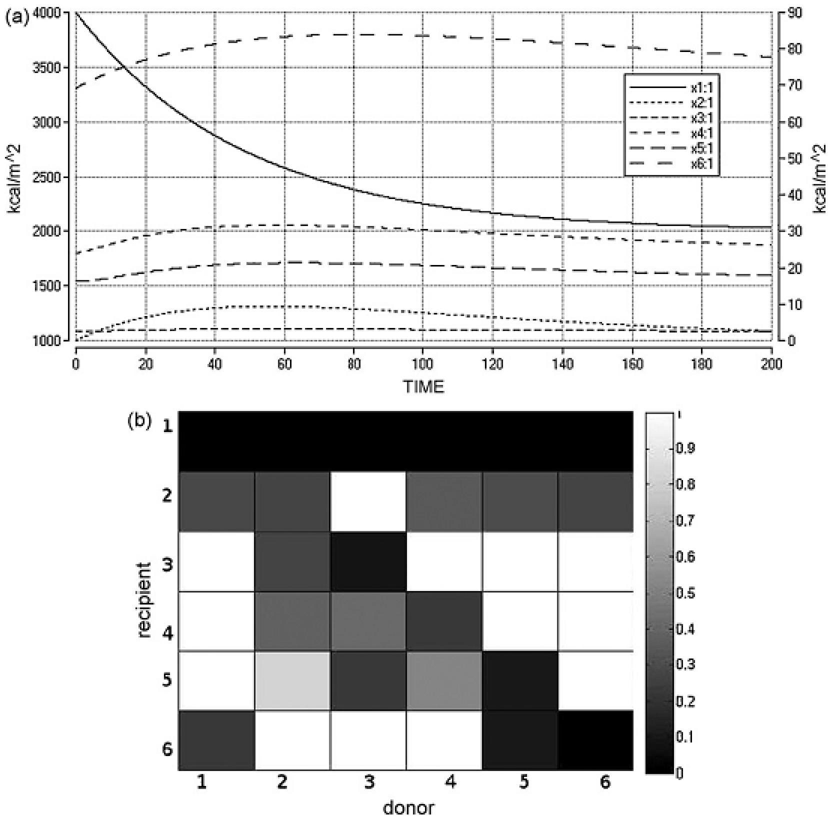


Figure 3.3: Evolution of stock values (a) and direct to indirect flow ratios (b) in the oyster model after the oyster compartment was doubled. x_1 and x_2 are on the left scale; all other stocks are on the right scale. Compartment abbreviations (x_1...x_6) are explained in Fig. 3.1.

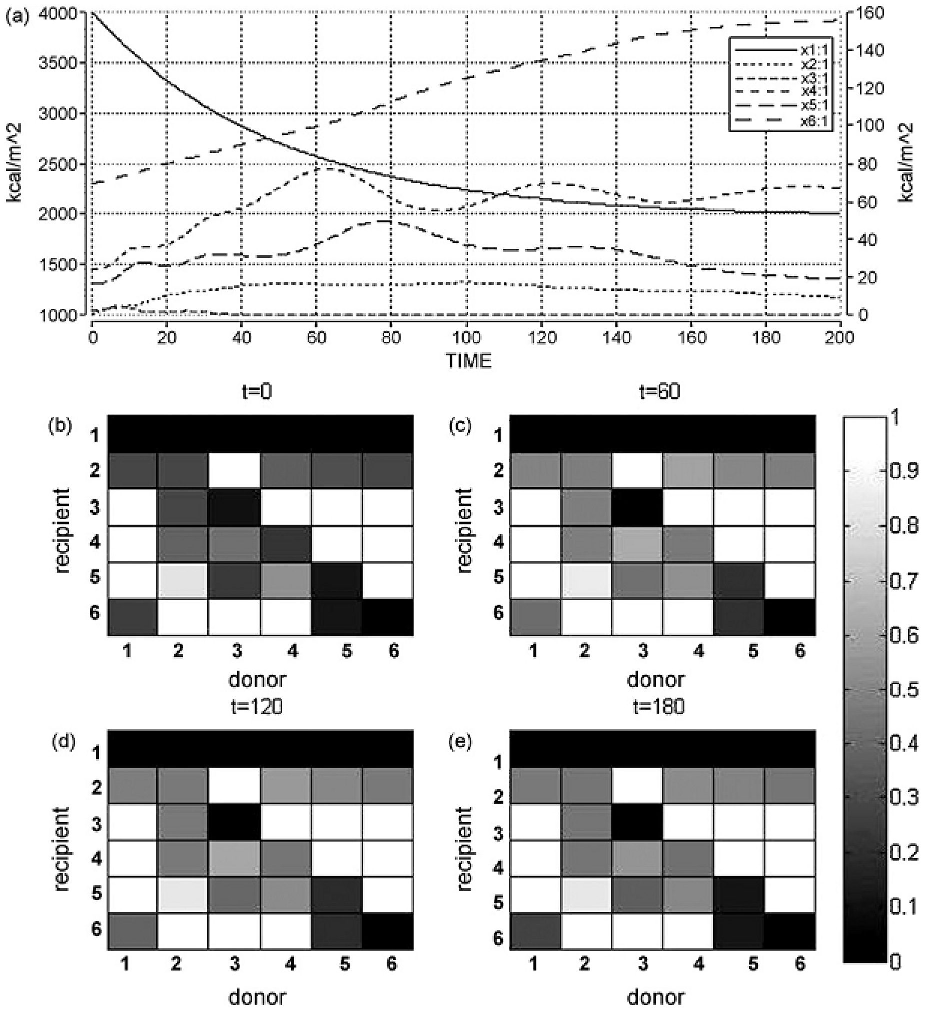


Figure 3.4: Evolution of stock values (a) and direct to indirect flow ratios (b–e) in a non-linear version of the oyster reef model (explained in text) after the oyster compartment was doubled. x_1 and x_2 are on the left scale; all other stocks are on the right scale. Compartment abbreviations ($x_1 ... x_6$) are explained in Fig. 3.1.

network model. Only output is used; what happens in the equations stays in the equations. This lets DEA sidestep the mathematical difficulties associated with prior methods.

Borrett et al. (2010) have found that, in empirically based trophic and biogeochemical models at a constant steady state, indirect effects become dominant after only a few terms of the infinite series expansion of the **N** matrix. These results are consistent with our finding that a window size of about 20 is sufficient to closely approximate **N**. We note that there might be better approximations than our method. However, the value of DEA methodology is in its intuitive definition. Eqs. 3.4 and 3.7 imply that any network characteristic investigated by NEA can also be studied with DEA, provided that the measure in question makes sense for a system away from steady state. It should be possible to investigate energy cycling (Patten, 1985) and system properties such as dominance of indirect effects (Patten, 1984), as well as the defined network properties of environs (e.g. Patten, 1995, Fath and Patten, 1999b). Other promising areas of application for dynamic environ approximation include the analysis of bioenergetic food web models, the study of system-level properties of individual-based models, including those incorporating evolution, and investigation of exact stochastic simulations of trophic dynamics. Applications such as these could provide a much-needed link between conventional and systems ecology.

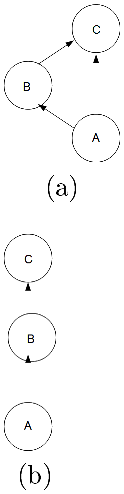
Chapter 4

Indirect Energy Flows in Niche Model Food Webs: Effects of Size and Connectance

4.1 Introduction

Food webs are icons of complexity, depicting intricate networks of feeding interactions. Since food webs can be studied both from the point of view of population dynamics and that of matter and energy flows, they bridge community and ecosystem ecology. Moreover, their study has led to insights that apply to other complex systems (McCann et al., 1998, Berlow, 1999, Csermely, 2006).

Examining food webs reveals a wide variety of indirect interactions, such as indirect matter and energy flows, trophic cascades, apparent competition, indirect mutualism and commensalism, and exploitative competition (Wootton, 1994). Indirect flows take place when energy or nutrients move between two species by a path, termed an indirect path, that includes one or more intermediate species (Figure 4.1). Previous work has shown that,



(a)

(b)

Figure 4.1: Direct and indirect flows. In the module shown in 4.1a, species A and C are linked by both direct and indirect flows, while in the module shown in 4.1b, they are linked exclusively by indirect flows.

although individual indirect flows may be small, their great number makes them important in ecosystems. In fact, in many empirically-based ecosystem models, the fraction of total energy flow that travels over indirect paths (flow indirectness or FI) is greater than 50%, a property often described as “dominance of indirect effects” (Higashi and Patten, 1986, Patten, 1986, Higashi and Patten, 1989, Fath and Patten, 1999b, Fath, 2004).

The mathematical and conceptual framework that allows flow indirectness and many other network properties to be quantified, termed environ analysis (Patten, 1978, Matis and Patten, 1981, Patten and Matis, 1982, Fath and Patten, 1999b), has not previously been applied to theoretical food web models with structures similar to those of field webs and empirically-based dynamics. Most studies of indirect matter and energy flows have focused on small, highly aggregated ecosystem models (Patten, 1985, Higashi and Patten, 1989, Fath, 2004), although some have looked at large, highly simplified, theoretical models (Fath, 2004) and steady-state empirical models of various sizes (Borrett et al., 2010).

This study investigates the importance of indirect energy flow in food webs by measuring the flow indirectness of theoretical food web models and examining how it is affected by web size and connectance, defined as the fraction of possible directed links that actually exist. Its purpose is not to provide a comprehensive examination of flow indirectness in various ecological models but simply to measure it in one commonly studied model and demonstrate the potential usefulness of environ analysis and DEA. The models studied here use the niche model (Williams and Martinez, 2000) for structure and the n-species Yodzis-Innes model (Yodzis and Innes, 1992, Williams et al., 2007) for dynamics. The niche model assumes that species in a community can be ordered along a “niche” dimension (Cohen, 1978), such that consumers mainly feed on species with a lower niche value than their own but may also feed on those with a higher niche value. The niche value is correlated with, but not identical to, body size (Williams and Martinez, 2000, Woodward et al., 2005, Williams et al., 2010). Each species feeds on all species (including, potentially, its own) whose niche value lies within a specified range. For species i with niche value n_i, the feeding range has width r_i and can be centered anywhere in the interval $[r_i/2,\ n_i]$ (Williams and Martinez, 2000).

The n-species Yodzis-Innes model uses consumer functional responses and the scaling of metabolic rate with body size (Schmidt-Nielsen, 1984, West et al., 1997, Brown et al., 2004) to add realism to a simple model of trophic dynamics. (Since, as described below, the analysis used to quantify flow indirectness requires a conservative currency, the model’s state variables were taken to be the total energy contents of each species.) Including a functional response that saturates at high prey density improves model realism by acknowledging the fact that there is a limit to how much food an individual can consume. The use of scaling relationships helps incorporate biologically reasonable sets of parameter values into a theoretical model. The model, which employs variables and parameters whose dimensions and values are listed in Table 4.1, is described below. In keeping with environ analysis convention, we consider energy to flow from column j to row i, not the other way around, as is the convention in dynamic food web modeling.

In the absence of consumers, producer j grows logistically at rate $r_j B_j(1 - \frac{B_j}{K_j})$, where B_j is the total energy content (or population size) of species j, r_j is its maximum growth rate and K_j is the environment's carrying capacity for species j. To obtain the rate at which species j is eaten by species i, we reason as follows. The rate of consumption of j by i is proportional to the population size of i, B_i. The quantity y_{ij} is the maximum rate at which species i can consume species j, divided by i's metabolic rate, x_i. Multiplying this quantity by x_i gives $x_i y_{ij}$, the maximum per-capita consumption rate for i preying on j. The functional response, $F_{ij}(B)$, gives the consumption rate as a fraction of this maximum rate, yielding $x_i y_{ij} F_{ij}(B) B_i$ for the actual rate. However, the predator does not ingest and assimilate all the prey it captures, so its consumption rate must increase to compensate for this. Dividing the previously obtained rate by the predator's efficiency, e_{ij}, accomplishes this, giving the expression $x_i y_{ij} F_{ij}(B) B_i / e_{ij}$.

We now turn to the functional response. Following Berlow et al. (2009), a sigmoidal (Holling Type III) functional response with predator interference (Skalski and Gilliam, 2001) was chosen, in part because it stabilizes the dynamics of food web models (Williams et al., 2006). In this model, the consumer's search rate is proportional to prey abundance raised to the power q, a non-negative real number (Gurney and Nisbet, 1998), and consumers of a given species interfere with each other with strength c (Skalski and Gilliam, 2001). As a result,

$$F_{ij}(B) = \frac{B_i^{1+q}}{B_0^{1+q} + cB_0^{1+q} B_j + \sum_{k=prey} B_k^{1+q}} \tag{4.1}$$

Following Berlow et al. (2009), the values $q = 1$ and $c = 1$ were used. (Table 4.1) The expression is dimensionless, as term by term analysis based on the dimensions listed in Table 4.1 reveals.

The overall differential equation for producer species j is:

$$\frac{dB_j}{dt} = r_j B_j (1 - \frac{B_j}{K_j}) - \sum_{i=predators} x_i y_{ij} B_i F_{ij}(B)/e_{ij} \tag{4.2}$$

Consumers of species i lose energy to metabolism at rate $x_i B_i$, gain it from prey item j at rate $x_i B_i y_{ij} F_{ij}(B)$, and lose it to consumers of species k at rate $x_k y_{ki} B_k F_{ki}(B)/e_{ki}$ (Williams et al., 2007). Overall, we have:

$$\frac{dB_i}{dt} = -x_i B_i + x_i B_i \sum_{j=prey} y_{ij} F_{ij}(B) - \sum_{k=predators} x_k y_{ki} B_k F_{ki}(B)/e_{ki} \tag{4.3}$$

Table 4.1 summarizes the model's parameters and their values.

To parametrize the model, empirically documented relationships between trophic level and body mass (Brose et al., 2006) were used to assign body masses to species in the model. Following Williams and Martinez (2004), the (usually non-integer) trophic level of each species was computed as the mean of two quantities: (1) the integer distance between the target species and the closest basal species (those that do not prey on any other species); and (2) Levine's (Levine, 1980) usually non-integer flow-based trophic position, computed under the assumption that predators receive equal fractions of their diet from all prey species. The equal flows assumption allows flow-based trophic positions to be assigned to species in a purely topological web. The expression for flow-based trophic position is:

$$TL_i = 1 + \sum_{j=1}^{S} TL_j p_{ji} \tag{4.4}$$

where TL_i is the trophic level of species i, S is the total number of species in the food web, and p_{ji} is the fraction of species i's diet provided by species j. The mean of this quantity and distance from a basal species was used because it can be computed from topological information and provides a close approximation to the true flow-based trophic position in food webs for which flow data are available (Williams and Martinez, 2004). Species were then assigned metabolic rates using the 3/4-power scaling relationship between metabolic rate and body size (Schmidt-Nielsen, 1984, West et al., 1997, Brown et al., 2004).

A new flow-based dynamic network analysis method called dynamic environ approximation (DEA; Shevtsov et al. (2009)) was used to compute FI. The basic logic of DEA is as follows. If a food web has adjacency matrix $\mathbf{A}$, then $\mathbf{A}^k$ gives the number of paths of length k between each pair of species and $\sum_{k=1}^{m} \mathbf{A}^k$ gives the total number of paths of length m or less between each such pair (Patten et al., 1976, Agnarsson and Greenlaw, 2006). If the structure of the network changed over time, then the number of paths would be given by the product series $\mathbf{A}(t)+\mathbf{A}(t)\mathbf{A}(t+1)+...+(\mathbf{A}(t)\mathbf{A}(t+1)...\mathbf{A}(t+m))$. DEA uses a related product series of matrices describing energy flows in the food web to trace the flows through the system. The flow matrix is then normalized by by the total outflow from the donor species to create a matrix, $\mathbf{G}$, of nondimensional flow intensities for each integer time step. Then, for a window of m time steps, we have integral flow $\mathbf{N}(t) = \mathbf{I}+\mathbf{G}(t)+\mathbf{G}(t)\mathbf{G}(t+1)+...+(\mathbf{G}(t)\mathbf{G}(t+1)...\mathbf{G}(t+m))$ and $FI_{ij}(t) = (N_{ij}(t) - G_{ij}(t))/N_{ij}(t)$ (Fath and Patten, 1999b, Shevtsov et al., 2009). This method was used to compute FI for each web and the entries of the $\mathbf{FI}(t)$ matrix were then averaged.

4.2 Methods

4.2.1 Web Construction and Simulation

The goal of this study was to explore the importance of indirectness in a commonly studied theoretical food web model, the niche model (Williams and Martinez, 2000). This model was selected because it is frequently studied and reproduces many features of real food webs with a fair degree of accuracy (Williams and Martinez, 2000). In this model, each species has a niche value, n_i, a feeding range width, r_i, that can be interpreted as the fraction of possible niche values that can be consumed by species i, and a feeding range center. The niche value

Quantity	Meaning	Dimensions (MLT)	Value used
B_i	energy content of compartment i	energy ($[M][L]^2[T]^{-2}$)	
r_j	intrinsic growth rate	1/time ($[T]^{-1}$)	1 for producers, 0 otherwise
K_j	carrying capacity for producers	energy ($[M][L]^2[T]^{-2}$)	1
x_i	body mass-specific metabolic rate relative to maximum producer growth rate	1/time ($[T]^{-1}$)	0.138 for producers, $0.314 m_i^{-1/4}$ otherwise
y_{ij}	consumption rate for i consuming j normalized by metabolic rate of i	none	8
e_{ij}	conversion efficiency for i consuming j	none	0.45 for feeding on producers, 0.85 otherwise
q	reward sensitivity in Holling Type III functional response	none	1
B_0	half-saturation density in functional response	energy ($[M][L]^2[T]^{-2}$)	0.5
c	strength of predator interference in functional response	1/energy ($[T]^2[M]^{-1}[L]^{-2}$)	1

Table 4.1: Parameters and variables of n-species Yodzis-Innes model where m_i is the body mass of species i. Parameter values were taken from Berlow et al. (2009).

for each species is drawn from a uniform distribution ranging from 0 to 1. Range centers are drawn from a uniform distribution ranging from $r_i/2$ to n_i. A uniform distribution is used both for its simplicity and to reflect the hypothesis that niche values in real ecosystems are roughly uniformly distributed.

Species' range widths are generated by drawing values from a beta distribution whose mean is twice the connectance of the web and multiplying them by the species' niche value. (The formula for the beta distribution is $p(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx}$ and the niche model uses $\alpha = 1$ and $\beta = \frac{1-2C}{2C}$, where $C = 2L/S^2$, the connectance of the web, and L is the number of links.) The beta distribution is used because it is a tractable distribution that ranges from 0 to 1 and whose mean can be defined by an input variable. Since the expected value of the beta distribution is $2C$ and that of n_i is 0.5, this procedure results in range width having an expected value of C. Because niche values are uniformly distributed on the [0,1] interval and a consumer's feeding range width is the fraction of this interval that contains potential prey, the fraction of species a given consumer preys on is approximately its range width. This gives the food web the desired connectance. Each species is assumed to prey on all species within its range, including itself, and a food web directed adjacency matrix is assembled.

Candidate webs generated by this method were tested to ensure that they had at least one producer and consisted of only one set of connected species, termed a *component* in graph theory (Agnarsson and Greenlaw, 2006). For the latter test, the Laplacian matrix, $\mathbf{L}$, was used. This matrix is defined as the difference between the degree matrix $\mathbf{D}$, which has the degree of the graph's nodes on the diagonal and zeros elsewhere, and the undirected adjacency matrix $\mathbf{A}$ (a symmetric matrix with $a_{ij} = a_{ji} = 1$ if a link exists between i and j and 0 otherwise), from which self-loops are excluded, making $a_{ii} = 0$. The equation for the Laplacian matrix is then $\mathbf{L} = \mathbf{D} - \mathbf{A}$. The number of times 0 appears as an eigenvalue of the Laplacian is the number of components in the graph (Anderson and Morley, 1971, Mohar, 1991).

If a web passed these tests, trophic levels were assigned to each species as the mean of distance from the closest basal species and the flow-based trophic position method (Eq. 4.4), computed under the assumption that predators receive equal fractions of their diet from all prey species (Levine, 1980, Williams and Martinez, 2004). (The equal flows assumption allows trophic levels to be computed for a purely topological web.) Taking the mean of these two methods used provides a good approximation to real trophic levels in quantitative food webs and ecosystem models (Williams and Martinez, 2004). Trophic levels were then used to assign body sizes as $10^{T_i - 1}$ where T_i is the trophic level of species i (Brose et al., 2006), and mass-specific metabolic rates were assigned using 3/4-power scaling (Schmidt-Nielsen, 1984, West et al., 1997, Brown et al., 2004). Initial abundances were drawn from a uniform distribution ranging from 0.5 to 1, ensuring that the simulation results were not artifacts of a particular set of initial conditions and that all species were initially present at ecologically significant levels. The simulation was then run for 1000 time steps using fourth-order Runge-Kutta integration with a step size of 0.01, after which time steady state had been reached or closely approximated. At that point, any extinct species were removed and the simulation run for 1000 more time steps. In order to avoid transient dynamics, only this second run was analyzed with DEA.

The effect of food web size (10 to 50 species) and connectance (0.1 to 0.48, in increments of 0.02) on flow indirectness was examined. Because the niche model is stochastic, 250 model realizations were generated and simulated for each pair of size and connectance values.

4.2.2 Dynamic Environ Approximation

In the standard Yodzis-Innes model, the amount of energy gained by a predator in a predation event is less than the amount lost by the prey. The boundary inputs and outputs required to balance the system's energy budget are not explicitly tracked. Therefore, in order to create the conservative flow matrix required for environ analysis, producer growth was conceptualized as an input to the system, while uneaten or unassimilated food and metabolic losses were conceptualized as outputs. For each integer time step, a flow matrix consisting of the second terms of Eq. 4.3 (with negative outflows from each compartment on the diagonal) was set and used to compute the throughflow-normalized flow matrix $\mathbf{G}$ ($g_{ii} = 0$, $g_{ij} = \frac{f_{ij}}{t_j^{out}}$, where f_{ij} is the energy flow from j to i and t_j^{out} is the total outflow from j). This was then used to perform DEA with a window size of 20, which previous work indicated is typically enough to capture all relevant dynamics (Shevtsov et al., 2009, Borrett et al., 2010). As a large majority of simulation runs had reached or nearly reached a constant steady state, the $\mathbf{N}$ matrix was only computed for one starting time. Flow indirectness (FI) was then calculated as $\mathbf{N} - \mathbf{G}$ and the mean for each web was computed. In only 62 runs out of 205,000, FI values larger than 1 or less than 0 were obtained; it was concluded that the integration step size was too large for the dynamics of these runs and they were excluded from further analysis. FI values for diagonal entries, which represent cycles linking a species to itself, were taken to be 0.

4.3 Results

Over the full range of parameter values, the mean flow indirectness of the model food webs was 0.092, with a standard deviation of 0.0279. It increased with system size but peaked at intermediate connectance levels, resulting in the pattern seen in Figure 4.2.

4.3.1 Determinants of flow indirectness

A classification and regression tree (CART) analysis performed in R (R Development Core Team, 2009) using the package rpart (Terry M Therneau and Beth Atkinson. R port by Brian Ripley., 2010) was used to explore which aspects of web structure were most strongly correlated with mean FI. Of 27 potential ecological and graph-theoretic predictor variables,

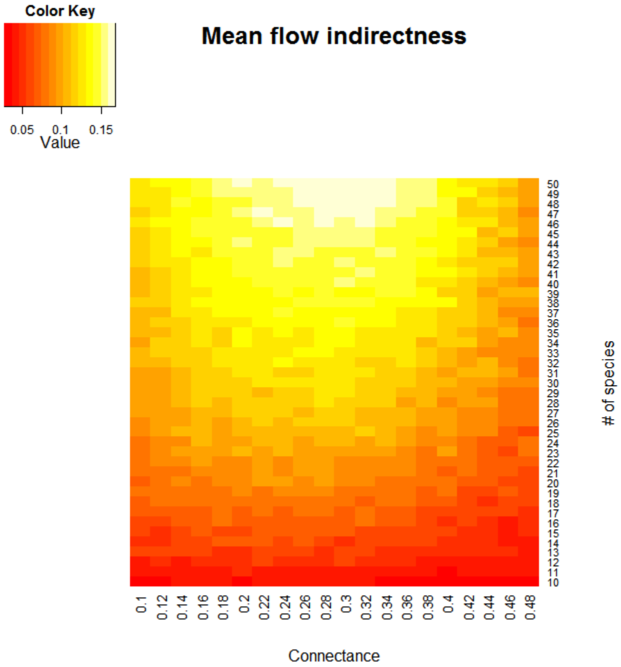


Figure 4.2: Mean flow indirectness as a function of nominal system size and connectance.

only five (mean path length, dominant eigenvalue of the adjacency matrix, connectance, mean trophic level and fraction of species belonging to intermediate trophic levels) were selected by the CART algorithm as best accounting for variation in the data. Two of these (mean path length and dominant eigenvalue of the adjacency matrix) were dominant. (Figure 4.3) The CART model accounted for 82.2% of the variation in FI. A full list and explanation of the potential predictor variables used is given in Appendix B.

4.3.2 Application to Field Data

The simulation results were compared to eight commonly studied topological food webs whose sizes and connectances fell within the range of the parameter scan. (References are in Table 4.2.) As Figure 4.4 shows, most of the relationships among food web size, connectance, mean path length and dominant eigenvalue in empirical food webs fall well within the range

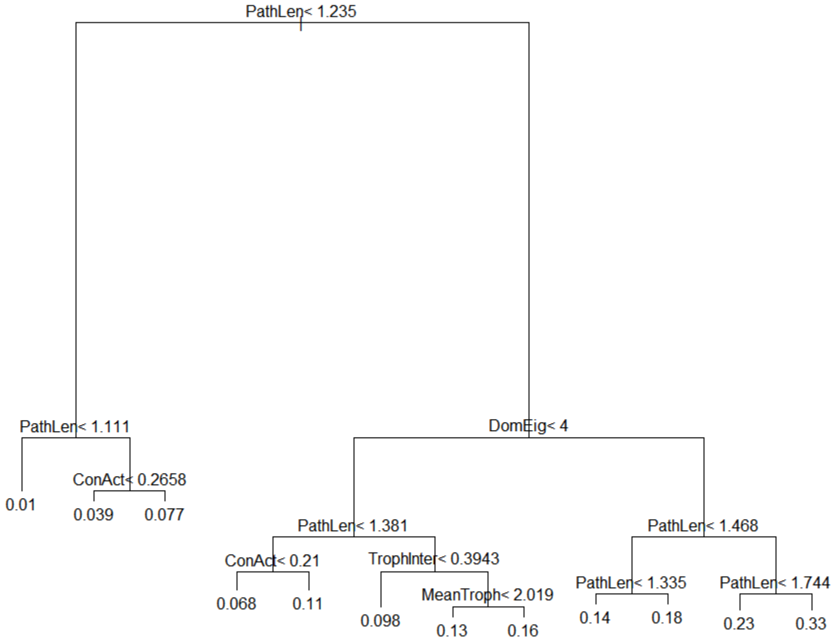


Figure 4.3: Classification and regression tree for flow indirectness. If the node condition is met, the left-hand branch is taken; otherwise, the right-hand branch is taken. Key to variables: PathLen - mean path length; DomEig - dominant eigenvalue of adjacency matrix; ConAct - actual connectance; TrophInter - fraction of species with both predators and prey; MeanTroph - mean trophic level.

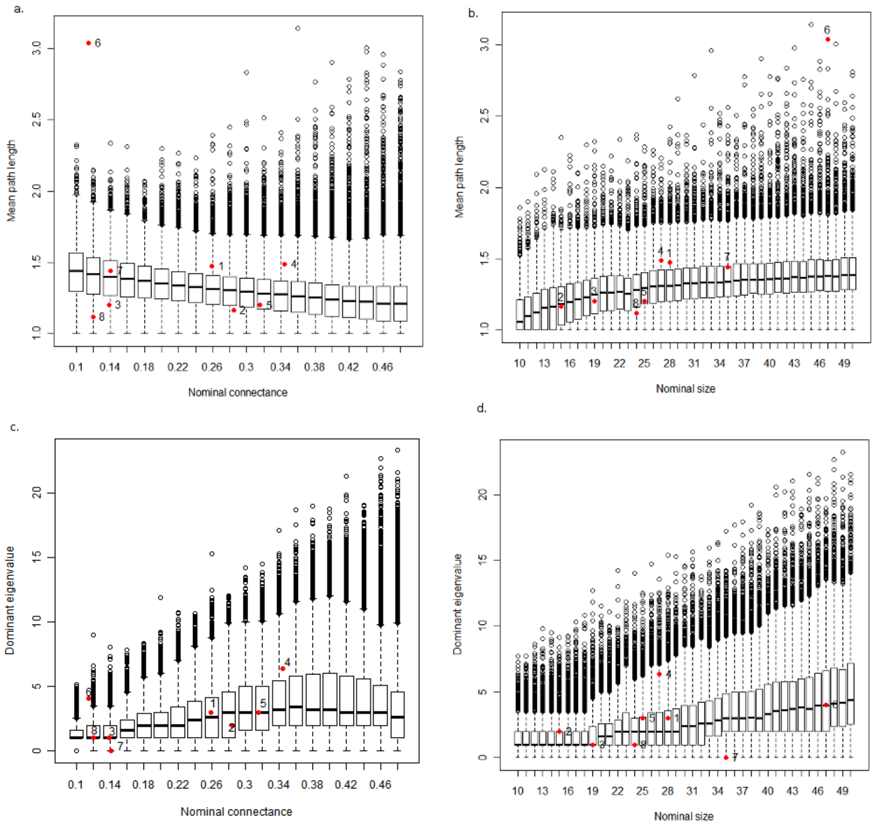


Figure 4.4: Model food web mean path length and dominant eigenvalue as a function of nominal size and connectance. Each boxplot shows the median, 75th and 25th percentiles for each data set, with whiskers extending up to 1.5 times the interquartile range and any points falling outside that area being plotted individually. In (a) and (c), mean path length generally decreases with connectance, while the dominant eigenvalue shows a curvilinear pattern. In (b) and (d), both mean path length and dominant eigenvalue increase with web size. Together, these relationships largely account for the patterns in Figure 4.2. Red numbered dots show where real food webs fall on these graphs. Key: 1. Benguela ecosystem; 2. Bridge Brook Lake; 3. Canton Creek; 4. Coachella Valley; 5. Skipwith Pond; 6. St. Marks Seagrass; 7. St. Martin Island; 8. Stony Stream

of model webs. The major exceptions are the St. Marks seagrass web, which has a much larger mean path length for its size and connectance than any of the other modeled or empirical webs, and the St. Martin Island web, whose dominant eigenvalue is zero because the web contains no cycles.

Predictions of the flow indirectness values of the empirical webs were made by following the CART tree developed using the simulated webs. (Table 4.2) Freshwater food webs had consistently lower predicted FI values than terrestrial and marine webs. This is due to the freshwater webs' lower mean path lengths, which appear to simply be a result of their smaller sizes. Size-normalized mean path length did not vary systematically with ecosystem type.

4.4 Discussion

Two major frameworks exist for studying networks of trophic interactions and the movement of energy within ecosystems: those of community and ecosystem ecology. When food webs are studied from a community ecology perspective, the emphasis is on individual species and their population dynamics. Such webs are as detailed as possible but often omit parts of the biota at the study location, especially decomposers and detritivores in terrestrial systems. (The desert food web of Polis (1991) is a prominent exception.) By contrast, the ecosystem framework uses comprehensive, usually highly aggregated models that focus on the movement of energy and nutrients. Researchers working within these two frameworks have ignored each other's work to a remarkable extent.

Environ analysis (Patten, 1978, Matis and Patten, 1981, Fath and Patten, 1999b) is a set of conceptual and mathematical tools for analyzing networks of stocks and flows. It has traditionally been applied to phenomenological models of real ecosystems. This is both a strength and a weakness, in that the analysis stays close to reality but is tied to a relatively small number of models that are usually highly aggregated. In particular, the six-

Ecosystem	Size	Connectance	Dominant eigenvalue	Mean path length	Intermediate fraction	Mean trophic level	Predicted FI
Freshwater							
Bridge Brook Lake (Havens, 1992)	15	0.284	2.00	1.16	0.93	2.35	0.077
Canton Creek (Townsend et al., 1998)	19	0.139	1.00	1.20	0.74	2.12	0.039
Skipwith Pond (Warren, 1989)	25	0.315	3.00	1.20	0.92	2.67	0.077
Stony Stream (Townsend et al., 1998)	24	0.12	1.00	1.12	0.75	2.25	0.039
Marine							
Benguela Current (Yodzis, 1998)	28	0.259	3.00	1.47	0.96	3.18	0.16
St. Marks Seagrass (Christian and Luczkovich, 1999)	47	0.115	4.03	3.04	0.87	3.52	0.33
Terrestrial							
Coachella Valley (Polis, 1991)	27	0.344	6.35	1.49	0.96	3.00	0.23
St. Martin Island (Goldwasser and Roughgarden, 1993)	35	0.140	0.00	1.44	0.77	2.62	0.16

Table 4.2: Network characteristics and predicted flow indirectness values for empirical food webs. Predictions were made by following the CART tree in Figure 4.3.

compartment intertidal oyster reef model of Dame and Patten (1981) may be the *Drosophila* of environ analysis because of the number of techniques and hypotheses that have been demonstrated and tested using it (e.g. Patten, 1985, Patten et al., 1990, Fath and Patten, 1999a, Whipple, 1999, Shevtsov et al., 2009).

This study is not the first to apply environ analysis to a large synthetic model. Fath (2004) created models of ecosystems by assigning species to one of six functional groups: primary producers, herbivores, carnivores, omnivores, detrital feeders and detritus. Each functional group, including detritus, contained the same number of species, ranging from five to one hundred. Biologically plausible intergroup interactions were then randomly assigned. The model used linear donor-controlled dynamics with randomly selected coefficients. Thus, this model possessed some realism with regard to functional groups, very little with regard to network structure, and almost none (except in the case of flows to detritus) with regard to dynamics.

The advantage of theoretical models such as the Yodzis-Innes model is that they describe causal relationships between species. Compartment models, on the other hand, are typically phenomenological, "bookkeeping" models. When dynamical assumptions such as donor control are added to these models, they are typically very simple and lack biological justification. The relatively detailed causal assumptions and parameter constraints of the Yodzis-Innes model may be criticized as being overly complex and unrealistic, but they are probably less wrong than linear models with donor control, which assumes that the amount of prey eaten by a predator species depends only on the prey's population size. However, research on such dynamically simple models has produced insights into ecosystem function and network properties.

Working within the framework of linear steady-state models of conservative energy and matter flows and storages, Patten et al. (1990) identified six network characteristics that directly increase flow indirectness in compartment models: number of compartments, connectance, storage, cycling, feedback and magnitude of direct flows. Most of the model webs in this study had much lower mean flow indirectness values than the ecosystem models examined in previous work (Fath, 2004, Borrett et al., 2010). This is likely due to the fact that niche model webs, unlike the models studied before, do not include detritus or detritivores. Therefore, they contain substantially less cycling than ecosystem-oriented models. Since cycling greatly increases the fraction of model currency traversing indirect paths (Patten et al., 1990), its absence must reduce flow indirectness. In nature, grazing- and detritus-based food webs intertwine (Odum, 1969, Wardle et al., 2004, 2005). Thus, the current results very likely underestimate the true importance of indirect flows in natural food webs and future work should attempt to include detritus and detritivores. Extending and modifying the niche model to incorporate both grazing- and detritus-based webs would represent an important advance.

Other important topics for future research include investigation of the sensitivity of these results to parameter values and model assumptions and the examination of energy cycling in model webs with and without detritus (Patten, 1985, 1986). In particular, the standard niche model uses niche values that are uniformly distributed between 0 and 1. However, the niche value is correlated with body size (Williams and Martinez, 2000, Woodward et al., 2005, Williams et al., 2010). Therefore, the distribution of niche values should be derived from body size distributions observed in nature. The allometric diet breadth model of Petchey et al. (2008) approaches this idea but relies on previously specified body size data, although this could be randomly generated. A simpler approach would retain all the assumptions of the niche model but use a more realistic distribution of niche values.

It will also be useful to examine the effects of other system attributes on the flow indirectness and find out whether this quantity is linked to the vulnerability of food webs to species loss. Dunne et al. (2002) found that, for sixteen empirical topological food webs, vulnerability to cascading extinctions in the face of species loss was negatively correlated with connectance and uncorrelated with the prevalence of omnivory (in spite of the correlation between omnivory and connectance). However, the webs examined in that study had connectances ranging from 0.026 to 0.315 – values falling within the range in which flow indirectness increases with connectance. (Figure 4.2) It would be instructive to examine the effects of species loss on model webs with higher connectance values, to see whether the positive relationship between connectance and robustness continues to hold and whether omnivory becomes a more important determinant of robustness as flow indirectness declines.

The results reported here help bridge contemporary community ecology and systems ecology, while providing a new way of looking at ecosystem complexity. It is also possible to apply dynamic environ approximation to non-trophic stock and flow networks such as dispersal networks (McRae and Beier, 2007) and human systems such as roads and economies, and doing so may provide insights into their function.

Chapter 5

A Gradient Analysis of Multiple Interacting Guilds in a Southern Appalachian Forest Highlights the Role of Biotic Interactions in Structuring Communities[1]

5.1 Introduction

The extent to which ecological communities are coherent entities as opposed to mere intersections of individual species distributions has long been one of the fundamental questions of ecology (Clements, 1936, Gleason, 1926, Tansley, 1935), to the point where studies addressing it are sometimes said to be about "the nature of the community" (Krebs, 2001). Gradient analysis, which goes back to Whittaker (1956), addresses this question by exam-

---

[1] Shevtsov, J. and K. Wickings. To be submitted to *Oikos*.

ining the degree of similarity of species distributions, particularly their endpoints. (Figure 5.1) However, all such studies have used organisms from a single taxon or guild. This unavoidably biases the results of these studies, as it risks missing important connections due to non-competitive interactions such as mutualism, predation (other than intraguild predation), parasitism, commensalism and amensalism (Putman, 1994). These types of interactions may be at least as important as competition in determining community structure and composition (Bruno et al., 2005, Stachowicz, 2001). Interactions beneficial to at least one party should be more likely to occur between members of different guilds. Such organisms are unlikely to compete for resources; however, they can have complementary niches that promote non-competitive interactions. Furthermore, resource competition tends to be diffuse in the sense that a species will compete with all other species sharing the same resource (Hubbell, 2001). By contrast, non-competitive interactions such as plant-pollinator, host-parasite, and some ectomycorrhizal relationships tend to be more specific (Bruno et al., 2005).

This study attempts to overcome the omissions of previous work described above by simultaneously examining the abundances of taxa belonging to four different guilds – seed plants, soil basidiomycetes, collembola and mesostigmatid mites – along an elevational gradient. These guilds were chosen because many basidiomycetes have either mycorrhizal or parasitic relationships with the plants, the diets of collembola include fungi, and mesostigmatid mites prey on collembola, along with other soil animals. We asked to what extent communities are coherent entities and tested the hypothesis that coherence is positively related to the number of interacting guilds studied. We also investigated whether interguild interactions are more common than intraguild ones.

To address these questions, we used two different methods of analyzing abundance and presence-absence data. The first, causal discovery, draws on work in computer science that allows the existence and directionality of some causal interactions to be inferred from correlational data (Pearl, 2009, Shipley, 2002, Spirtes et al., 2001). The second, Elements of

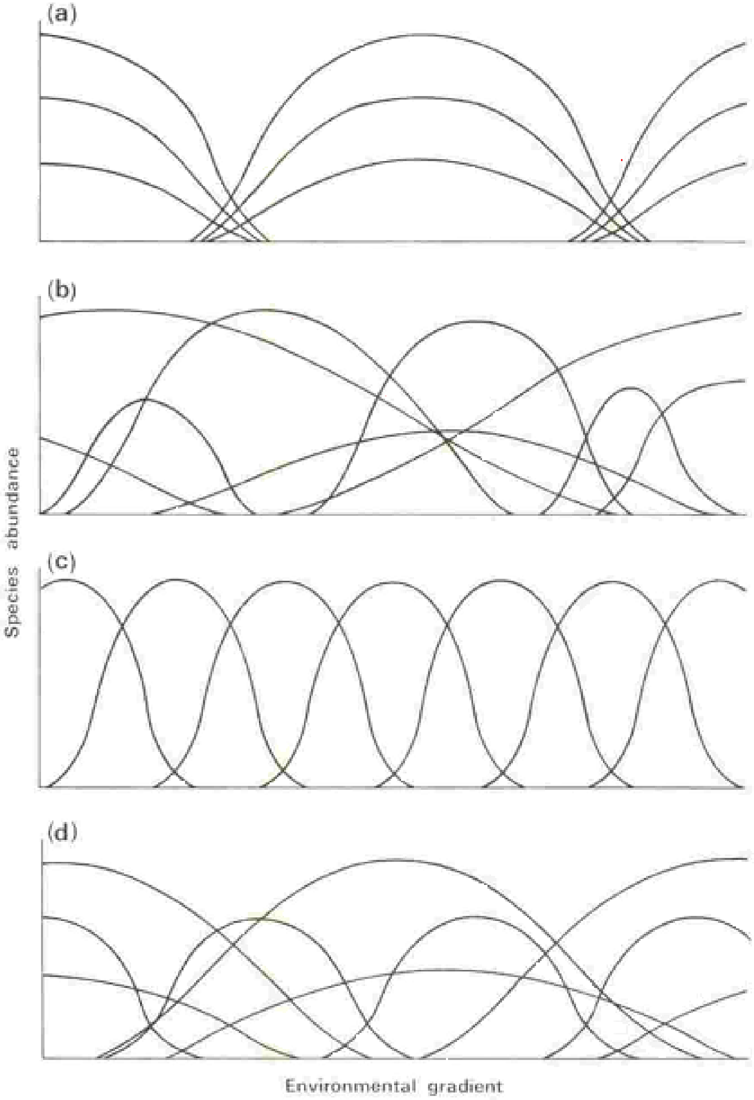


Figure 5.1: Four possible patterns of vegetation distribution along a gradient: (a) a tightly integrated community; (b) an individualistic community; (c) a community structured by resource partitioning; (d) a community of several guilds, each of which is structured by resource partitioning. From Austin (1985). EMS analysis, summarized in Fig. 5.3, would classify community a as Clementsian, community b as Gleasonian, community c as evenly spaced, and community d as either Clementsian or Gleasonian, depending on the amount of boundary conjunction.

Metacommunity Structure, looks at community coherence, species turnover and boundary conjunction in order to classify communities (Leibold and Mikkelson, 2002, Presley et al., 2010a).

The term "metacommunity" refers to a set of local communities that may be linked by dispersal (Leibold and Mikkelson, 2002). However, in this study and much prior work (Whittaker, 1956, 1960, Pielou and Routledge, 1976, Bossenbroek et al., 2005), the local sites are merely samples of a continuously vegetated landscape. In order to link our work to previous research on community structure, we will use the term "community" rather than "metacommunity" throughout this article.

Elements of Metacommunity Structure (EMS) analysis examines community coherence, turnover and boundary conjunction in order to classify communities by the spatial patterns of taxon co-occurrence they exhibit. In a highly coherent community, there are relatively few gaps in taxon distributions, which is likely to occur when the abundances of different taxa respond to the same environmental gradient or gradients (Presley et al., 2010a). If not all small ranges are embedded within larger ones following a strict hierarchy, some taxa are replaced by others over the length of the gradient. The opposite of high replacement is high nestedness (Leibold and Mikkelson, 2002). (Figure 5.2) Finally, boundary conjunction refers to the frequency with which the first or last occurrences of some taxa along a gradient coincide with those of others, with high boundary conjunction indicating more discrete communities (Leibold and Mikkelson, 2002, Presley et al., 2010a). Communities are classified based on the combination of coherence, turnover and boundary conjunction values they exhibit. (Figure 5.3)

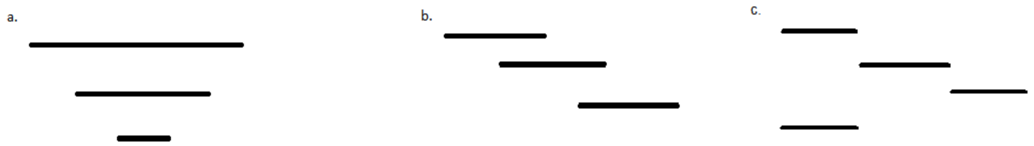


Figure 5.2: Relationship between nestedness and replacement. In each panel, an environmental gradient runs from left to right and the lines represent species' ranges. In Fig. 5.2a, nestedness is maximal and replacement is zero; in the other two panels, replacement is maximal and nestedness is zero. Fig. 5.2c, boundary conjunction is maximal; in the other panels, it is zero. All ranges portrayed exhibit maximal coherence, which would be reduced by the introduction of gaps.

5.2 Methods

5.2.1 Site Description

Coweeta Hydrological Laboratory is located in Macon County, North Carolina, USA, in the Blue Ridge physiographic province (Hunt, 1973). The bedrock is granite and Carolina gneiss (Boring et al., 1981). Precipitation averages 200 cm per year, increasing with elevation; the average temperature minimum is -17ºC and the average maximum is 33ºC. The southern Appalachians were not glaciated during the last ice age; however, soils at high elevation are immature. The soils at our study sites are various sandy loams (Soil Survey Staff, Natural Resources Conservation Service, 2012); data on soil chemistry is in Table 5.1. The forests are largely second growth, with oak (*Quercus* spp.), hickory (*Carya* spp.), tulip poplar (*Liriodendron tulipifera*) and rhododendron (*Rhododendron* spp.) being important. The land has not been logged since 1923 and grazing was stopped in 1933 (Douglass and Swank, 1975, Nelson, 1955). Chestnut blight (*Cryphonectria parasitica*) largely eliminated the American chestnut (*Castanea dentata*) from these forests between the 1930s and 1950s (Nelson, 1955), although some *C. dentata* sprouts were observed. Today, they typically grow to sapling height but no more. In addition, the hemlock wooly adelgid (*Adelges tsugae*) is greatly reducing the abundance of eastern hemlock (*Tsuga canadensis*) (Krapfl et al., 2011).

Six sites spaced at roughly equal elevational intervals were set up along an elevation gradient (685 m to 1520 m). Each site consisted of five 20×20 m haphazardly located plots, typically clustered near a small, unpaved road or trail. Plots with missing data for one or more guilds were omitted from the analysis, resulting in four plots from each site being used in the analysis, with the exception of Site B (1280 m), for which all five plots were used.

5.2.2 Plant Survey

In each 20×20 m plot, all living trees with diameter at breast height (DBH, 1.3 m) of at least 5 cm were censused and their DBH was measured. Shrub and herb cover was measured by surrounding each plant or cluster of plants with a triangle or rectangle and recording the dimensions of each polygon. Shrub cover was measured in full plots, while for herbs, a 5×5 m subplot was set up near the center of each full plot. Vines were omitted due to a lack of satisfactory methods for quantifying their abundance.

5.2.3 Soil Chemistry, Fauna and Fungi

Five randomly placed soil samples were collected from each plot from 0-5 cm in depth. This zone was chosen because it exhibits high biological activity and is commonly used in soil faunal surveys (Coleman et al., 2004).

5.2.3.1 Soil Chemistry

For chemical analysis, a composite sample was made for each site by combining equal volumes of soil from each sample in each plot at the site. The composite samples were analyzed for C, N, P, K, Ca, Na, Mg and pH at the Odum School of Ecology Analytical Laboratory. For all analyses except pH, the soil was oven-dried for 24 hours at 80$^{\underline{o}}$C and ground to 250

μm or finer using a ball mill. C and N were assayed by Micro-Dumas combustion (Kirsten, 1983), and total P was measured using an acid persulfate digest (Nelson, 1987). For K, Ca, Na and Mg, a double acid extraction was performed and concentrations were measured by atomic absorption spectrophotometry. pH was measured using the method described in (Peech, 1965).

5.2.3.2 Fungi

Basidiomycete operational taxonomic units (OTUs) were defined and their relative abundance quantified using terminal restriction fragment length polymorphism analysis (T-RFLP) (Dickie and FitzJohn, 2007). For molecular analysis, DNA was extracted from each sample using a MoBio UltraClean Soil DNA Isolation Kit according to the manufacturer's instructions. PCR was performed using the ITS-1F (specific to fungi) and ITS-4B (specific to basidiomycetes) primers (Gardes and Bruns, 1993). The PCR mix consisted of 2 μL of template DNA and 48 μL of a mixture consisting of 1× PCR Buffer (ABI), 2mM $MgCl_2$, 0.2 mM each deoxynucleoside triphosphate, 1μM forward primer (ITS1-F), 1μM 5'-FAM-labeled reverse primer (ITS4-B), 0.4 μg/μL bovine serum albumin, and 0.025 U/μL DNA polymerase. An MJ Research PTC 200 thermocycler was used for amplification with the following thermocycling pattern: an initial denaturation step consisting of 5 min at 94^{o}C; 35 cycles consisting of 0.5 min at 94^{o}C, 0.5 min at 52^{o}C and 1 min at 72^{o}C; and a final elongation step of 5 min at 72^{o}C (Edwards et al., 2004, Landeweert et al., 2005).

The PCR products were digested at 37^{o}C for 20 minutes with EcoRV and HaeIII restriction enzymes. Cleanup was immediately performed using the QiaQuick Nucleotide Removal Kit (Qiagen). Then, T-RFLP was performed using an ABI 3130 Genetic Analyzer with 14.25\mu L of HiDi Formamide, 0.75μL of ILS600 as an internal lane standard, and 5μL of purified digest (Dickie and FitzJohn, 2007).

5.2.3.3 Soil Fauna

In order to identify and count collembolans and mesostigmatid mites, soil samples were placed in a Berlese funnel for three days, with collected fauna being preserved in 70% ethanol. The collected animals were sorted and identified using a stereo microscope connected to a television screen. Collembolans were identified to family under a stereo microscope following Triplehorn and Johnson (2005). Identification of mesostigmatid mites was conducted on slide mounted specimens using a compound microscope and keys provided by the Ohio State Soil Acarology Summer Program along with an interactive computer key to Mesostigmata (Lucid Player Standard v2.2).

Data Analysis

Elements of Metacommunity Structure

Elements of Metacommunity Structure (EMS) analysis examines community coherence, species replacement and boundary conjunction. First, an ordination is performed on the presence-absence matrix and the sites and species (rows and columns of the presence-absence matrix) are sorted by their scores on the first axis given by the ordination, giving what is called the ordinated matrix. Most EMS studies use correspondence analysis, an ordination method that minimizes the distance between similar sites and similar species in a postulated environmental space (Heino, 2005, 2009, Leibold and Mikkelson, 2002, Presley et al., 2009, 2010b). However, in order to specifically examine responses to changes in elevation, canonical correspondence analysis with elevation as the constraint (Quinn and Keough, 2002, Heino, 2005) was performed on the presence-absence matrix. In canonical correspondence analysis, the axes of the environmental space must be as correlated as possible with linear combinations of specified environmental variables (Quinn and Keough, 2002).

Having obtained the ordinated matrix, the number of embedded absences in the range of each taxon was used as a measure of coherence, with many absences signifying low coherence (Leibold and Mikkelson, 2002). In order to obtain a measure of statistical significance, these results were compared to those from a null model in which taxa were randomly assigned to the same number of plots they were found in in real life, independently of each other and elevation. Thus, commonness and rarity were preserved, but the number of taxa in each plot was free to vary. This null model was chosen because all survey plots were of the same size and plots at a given elevation were similar enough that there did not appear to be strong reasons why some should have a higher taxonomic richness than others. Taxonomic replacement was quantified as the number of times the ranges of two taxa, A and B, overlapped in such a way that A's range started and ended before B's. (Figure 5.2b) When this happens, we say that A is replaced by B over the length of the gradient. The number of such replacements was compared to those from the null model just described (Hoagland and Collins, 1997, Leibold and Mikkelson, 2002). Range boundary conjunction was measured using Morisita's Index:

$$I = Q\sum_{i=1}^{Q}\frac{n_i(n_i-1)}{N(N-1)}$$

in which Q is the total number of plots in the transect, n_i is the number of taxon range boundaries (first or last occurrences of a taxon) in the i'th plot, and N is the total number of boundaries. If Morisita's Index is equal to one, range boundaries are distributed independently of each other, if it is greater than one, they are clumped, and if it is less than one, they are hyperdispersed or evenly spaced. A chi-square test was used to determine statistical significance (Hoagland and Collins, 1997, Leibold and Mikkelson, 2002). In order to ensure comparability with prior work, a significance level of 0.05 was used.

Communities were classified according to the system laid out in Figure 5.3 and described in detail in Presley et al. (2010a). In this system, both Clementsian and Gleasonian communities have high range coherence and replacement, as taxon abundances are driven by

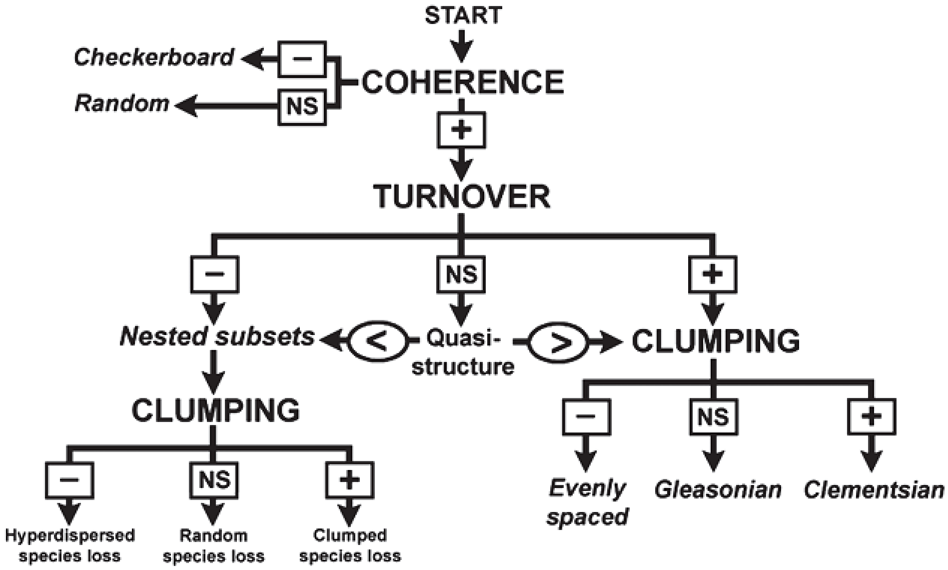


Figure 5.3: A flow chart illustrating the classification of communities by coherence, turnover (replacement) and conjunction, modified from Presley et al. (2010a). "+" and "-" symbols indicate statistically significant increases or decreases in the quantity in question compared to a null model. "NS" indicates lack of statistical significance; ovals show directionality of non-significant results. Note that Clementsian and Gleasonian metacommunities are distinguished only by the presence or absence of statistically significant boundary conjunction.

environmental gradients in both cases. The presence of significant amounts of boundary conjunction in Clementsian communities corresponds to Clements' tightly integrated community concept, while the absence of significant boundary conjunction in communities classified as Gleasonian comes from Gleason's conception of species abundances as independent of each other (Leibold and Mikkelson, 2002). Quasi-structures are distinguished from regular ones by the absence of statistically significant increases or reductions in taxonomic replacement. For example, a quasi-Clementsian community has significantly elevated coherence, high but not significantly elevated replacement, and significantly elevated boundary conjunction.

Causal Discovery

Causal discovery algorithms were developed by computer scientists and allow some causal relationships to be inferred from observational data. The following discussion draws heavily on Shipley (2002) and Pearl (2009).

Causal discovery methods rest on the notion of conditional independence. Two variables, X and Y, are conditionally independent given a set of variables $\mathbf{Z}$ (not including X and Y themselves) if knowing the values of X (or Y) and $\mathbf{Z}$ gives us no more information about Y (or X) than knowing $\mathbf{Z}$ alone. Many causal structures can be distinguished from one another by the conditional independencies they entail.

Consider three variables, A, B, and C, whose causal relationships are shown in Figure 5.4a. Measuring their values would reveal a correlation between A and C; this correlation arises because A is an indirect cause of C. However, we can account for the influence of B (a process known as *conditioning* on B) by, say, performing a regression with B as the independent variable and C as the dependent one. If we condition on B before testing for correlation, the association between A and C vanishes, as the causal influence of A on C is entirely mediated by B. Thus, A and C are conditionally independent given B.

On the other hand, in Figure 5.4b, termed a collider, A and B are uncorrelated (ignoring the possibility of an accidental correlation) but become correlated when C is taken into account. To understand this phenomenon, called Berkson's paradox, an example offered by Pearl (2009) is useful. Consider a college that requires either a high SAT score (A) or athletic ability (B) for admission (C). Since strong athletic ability can compensate for a low SAT score and vice versa, the two quantities will be negatively correlated among students who have been admitted to the college, *even if they are independent in the larger population*. Therefore, we can distinguish the two causal structures in Figure 5.4 by the patterns of conditional independence relationships they entail – the causal chain in Figure 5.4a produces a situation where A and C are correlated but become uncorrelated given B,

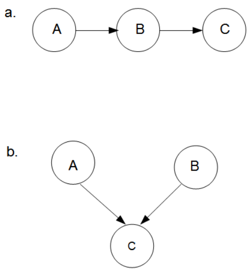


Figure 5.4: Examples of causal structures. (a) causal chain; (b) collider

while the collider in Figure 5.4b makes A and B, which are independent if measured on their own, correlated given C.

The goal of causal discovery is to produce a directed graph (typically acyclic) representing direct causal relationships between the members of a set of variables. (In practice, it is rarely possible to pick out a single graph and we instead find a family of causal graphs, termed observationally equivalent causal graphs, with the same undirected skeleton and set of forks and colliders.) In order to do this, it is necessary to make some assumptions about the causal structure we are looking for. The first assumption, called *stability* or *faithfulness*, simply states that observed patterns of conditional independence are due to the structure of the causal network rather than specific parameter values. Consider a photograph of a tree. It could be showing one tree, or there could be a second tree directly behind the first. One of the reasons we prefer the first explanation is stability, as only a very specific camera angle will make two trees look like one (Pearl, 2009).

The second assumption is a version of Occam's razor, in which a simpler model is preferred to a more complex one. In order to understand this assumption, we must recognize that we are very rarely able to measure all relevant causes of a set of variables. The unmeasured variables, termed *latent variables*, may be included in a proposed causal structure. This allows an infinity of possible causal structures to be created, so we must select the simplest ones (Spirtes et al., 1995, Pearl, 2009). However, here simplicity is defined not in terms of the number of links in a causal structure but in the ability of one structure to mimic the patterns of conditional independencies entailed by another. If two causal structures are consistent with the same data, we prefer the one that is compatible with a narrower range of observed values of the variables (Pearl, 2009). Assuming that the causal structure being searched for is minimal and stable allows the use of causal discovery algorithms. Most such algorithms, including the one used in this paper, also assume that the causal structure is acyclic. Algorithms that can recover cyclic structures exist, but the data must be discrete or meet certain distributional assumptions, or the functional relationships between variables must be linear.

We used the Fast Causal Inference (FCI) algorithm (Spirtes et al., 1995) to search for causal relationships between the abundances of taxa. This algorithm was chosen because it allows for latent variables, can handle continuous data and does not require the data to follow any particular distribution.

FCI first creates a skeleton of undirected links and then orients as many of them as possible from cause to effect. It starts by building a complete undirected graph of variables (in this case, taxa) and then picks pairs of variables and searches for conditioning variables that make the members of the pair conditionally independent. If it is possible to make a pair of variables conditionally independent, the link is removed. The remaining links are then oriented according to the rules described in Spirtes et al. (1995).

The abundance values for each taxon were normalized by the taxon's mean abundance. A Gaussian distribution, chosen as a model of a humped response because of the poor performance of beta distributions in modeling species response curves (Oksanen and Minchin, 2002), was fitted to each taxon's normalized abundance values to control for elevation; the residuals were used in further analysis. Standard minor axis regression, which is appropriate when the predictor variables are uncontrolled and may even be affected by the dependent variables (Quinn and Keough, 2002, McArdle, 1988, 2003), was then applied to the residuals to control for soil carbon, nitrogen, phosphorus, calcium, magnesium and potassium concentrations. (Sodium and pH were omitted due to their collinearity with other predictor variables.)

The Fast Causal Inference algorithm was then applied to the data using the R package pcalg (Kalisch et al., 2011), with conservative skeleton construction. The absolute value of the Spearman correlation coefficient for two variables, with a threshold of 0.5, chosen to be moderately high, was used as the test of independence; Spearman partial correlation was used for conditional independence. We chose to use the absolute value of the correlation rather than a p-value to test independence because correlation is more scientifically meaningful as a measure of association strength than p-values, which are strongly influenced by sample size (Ziliak and McCloskey, 2008).

In order to determine whether inter- or intra-guild interactions were more common, all unidirectional and partially oriented links found by FCI were classified according to the guild memberships of the taxa involved and the frequency of each type of interaction was compared to the values expected if all interaction types were equally probable. Since the sample sizes in such a comparison are necessarily small (n=6 and n=4), it was not possible to use statistical tests based on the Gaussian distribution on our data, as it is impossible to establish that 4 or 6 points have a Gaussian distribution. Therefore, a non-parametric, bootstrap two-group ranked comparison of medians with 10,000 bootstrap replicates (Good, 2005) was used to test for statistical significance.

Site	Elevation (m)	C (%)	N (%)	Total P (%)	Ca (%)	Mg (%)	Na(%)	K(%)	pH
A	1520	16.76	0.97	0.1130	0.0012	0.0010	0.0017	0.0241	5.00
B	1280	8.703	0.49	0.0663	0.0023	0.0009	0.0021	0.0240	5.61
C	1173	21.55	0.807	0.0265	0.0031	0.0008	0.0069	0.0379	4.99
D	968	9.0	0.56	0.0524	0.1300	0.0007	0.0026	0.0404	6.31
E	845	9.21	0.393	0.0446	0.0015	0.0010	0.0022	0.0253	5.73
F	687	3.72	0.22	0.0389	0.0037	0.0009	0.0028	0.0221	5.56

Table 5.1: Site elevation and soil chemistry. C and N values are means of three replicate measurements. All other chemistry values are for composite samples from n=4 plots for all sites except B, where n=5 plots. The soil at Sites B, C and F is a fine sandy loam, that at Sites D and E is a cobbly sandy clay loam, and that at Site A is a sandy loam (Soil Survey Staff, Natural Resources Conservation Service, 2012).

5.3 Results

5.3.1 Soil Chemistry

The soil chemistry results are shown in Table 5.1. Site D had significantly higher soil pH and calcium concentration than the other sites, while Site A is distinguished by its low pH and high soil carbon and phosphorus. Site C also has relatively acidic, carbon-rich soil, but its phosphorus level is the lowest of all the sites.

5.3.2 Elements of Metacommunity Structure

The results of EMS analysis are shown in Table 5.2. The overall subcommunity was classified as Clementsian, plants and mesostigmatid mites are quasi-Clementsian, basidiomycete fungi are Clementsian, and collembolans are random.

As indicated by the highly significant coherence values in Table 5.2, the overall subcommunity showed strong coherence along the elevational gradient, as did every guild except collembolans. Also, in all cases, Morisita's index (MI) was substantially greater than one, indicating that species range boundaries tended to coincide. Replacement values were higher than the simulated null for the full community and all taxa except collembola, for which they were much lower than the null. (Table 5.2) Removing Site C, which was very different from all the other sites in terms of both vegetation and soil chemistry (Table 5.1), from the analysis only changed the classification of the basidiomycetes, which went from Clementsian to Gleasonian. (Since Morisita's index for the basidiomycetes is the least significant of any guild, the change in their classification with the removal of Site C is likely to be partly due to the change in sample size.) This implies that our results are largely robust with respect to the presence of an unusual site.

In order to make sure that the Clementsian and quasi-Clementsian results were not likely to be artifacts of low sampling resolution, Morisita's index was computed for the same null model used to evaluate coherence and replacement. Out of 1000 runs, only five had MI values as large as the one observed for the full subcommunity, indicating that our sampling resolution does not create spurious boundary conjunction when species are distributed independently. As another check, we examined Whittaker's Smoky Mountains tree data for mesic sites (Whittaker, 1956), which was one of the examples in Leibold and Mikkelson (2002). That study used correspondence analysis and failed to obtain an MI value that would indicate statistically significant boundary conjunction. We used canonical correspondence analysis with elevation as the constraint to ordinate the data matrix and obtained the same result. To test the effect of sampling resolution, five of Whittaker's eleven sites were then discarded, leaving six sites at 800 ft intervals instead of eleven sites at 400 ft intervals. This actually made Morisita's index smaller, going from 1.1684 to 0.8222. These two analyses together indicate that the results presented here are unlikely to be an artifact of low sampling resolution.

In order to examine the effect of interguild interactions on taxon distributions, we performed EMS analysis on pairs and triplets of guilds, as well as the full subcommunity (Table

		Coherence				Replacement				Boundary conjunction		
Level	*Taxon*	*Abs*	*P*	*Mean*	*SD*	*Rep*	*P*	*Mean*	*SD*	*Morisita's Index*	*P*	*Metacommunity structure*
Single	Plants	802	0.0026	942.3	46.5	61366	0.7555	57661	11899	4.0444	<0.0001	quasi-Clementsian
	Fungi	1163	0.0002	1407	68.9	103185	0.0006	55775	13761	1.5523	0.0419	Clementsian
	Collembola	25	0.3429	19.45	5.85	33	<0.0001	268.6	57.5	1.5026	<0.0001	Random
	Mesostigmatids	75	0.0011	117.6	13.1	1618	0.2810	1358.7	240.6	1.4526	0.0021	quasi-Clementsian
Double	P + F	1792	0.0006	2011.9	63.8	123479	0.0008	67052	16888	2.6341	0.0001	Clementsian
	P + C	562	<0.0001	716	33.4	21057	0.0574	14228	3594	2.8455	<0.0001	quasi-Clementsian
	P + M	621	<0.0001	845.6	35.1	35505	0.0019	20271	4893	2.6481	<0.0001	Clementsian
	F + C	1097	0.0073	1245.6	55.4	58575	0.0063	34373	88656	1.5737	0.0315	Clementsian
	F + M	1190	0.0008	1383.4	57.7	86304	0.0025	48640	12460	0.8990	0.3858	Gleasonian
	C + M	107	<0.0001	161.2	12.8	1878	0.3815	1556	368.4	1.6000	0.0004	quasi-Clementsian
Triple	P + F + C	1861	0.0014	2060.4	62.4	116651	0.0014	64718	16228	2.5668	0.0002	Clementsian
	P + F + M	1958	0.0002	2201.4	64.5	146261	<0.0001	73354	18191	1.9723	0.0106	Clementsian
	P + M + C	682	<0.0001	896.4	34.6	34095	0.0014	19662	4525	2.8339	<0.0001	Clementsian
	F + C + M	1283	0.0038	1449.6	57.5	87524	0.0035	49008	13200	0.5636	0.0697	Gleasonian
Full	P + F + C + M	2712	0.0007	2943.5	90.3	350416	0.0127	224310	50626	2.4947	0.0007	Clementsian

Table 5.2: EMS analysis. "Abs" means absences within a taxon's range and "Rep" is replacements along the gradient. Means and standard deviations are from the null model simulation. Communities were classified according to the scheme in Figure 5.3. P = Plants, F = Fungi, C = Collembola, M = Mesostigmatids.

Level	Guilds	Observed MI	Baseline MI
Double	Plants + fungi	2.6341	2.4868
	Plants + collembola	2.8455	3.8489
	Plants + mesostigmatids	2.6481	3.5612
	Fungi + collembola	1.5737	1.5499
	Fungi + mesostigmatids	0.8990	1.5402
	Collembola + mesostigmatids	1.6000	1.4659
Triple	Plants + fungi + collembola	2.5668	2.4570
	Plants + fungi + mesostigmatids	1.9723	2.4050
	Plants + collembola + mesostigmatids	2.8339	3.4305
	Fungi + collembola + mesostigmatids	0.5636	1.5387
Full	Plants + fungi + collembola + mesostigmatids	2.4947	2.3797

Table 5.3: Comparison of actual and baseline Morisita's index (MI) values for sets of multiple guilds. The baseline MI value for a set of guilds is the weighted mean of the MI values of the individual guilds. Only sets of directly interacting guilds (six out of eleven sets in total) have higher-than-baseline boundary conjunction values, and such boundary conjunction values are found in all sets of directly interacting guilds except the fungi, collembola and mesostigmatids triplet.

5.2). All pairs except fungi and mesostigmatids had a Clementsian or quasi-Clementsian structure, as did all triplets except fungi, collembola and mesostigmatids, and the full subcommunity. The values of Morisita's Index for these sets of guilds were then compared to baseline values (Table 5.3). The baseline value for a pair or triplet was computed as the weighted mean of the MI values of the guilds composing it. The observed MI value was greater than the baseline one (and, in one case, greater than both values for the pair's component guilds) for the plants-fungi, fungi-collembola and collembola-mesostigmatid pairs. In the case of triplets, only plants-fungi-collembola had a greater than baseline MI, as did the full subcommunity. This creates a striking pattern in which only sets of directly interacting guilds have higher-than-baseline boundary conjunction values, and such boundary conjunction values are found in all but one set of directly interacting guilds (Table 5.3).

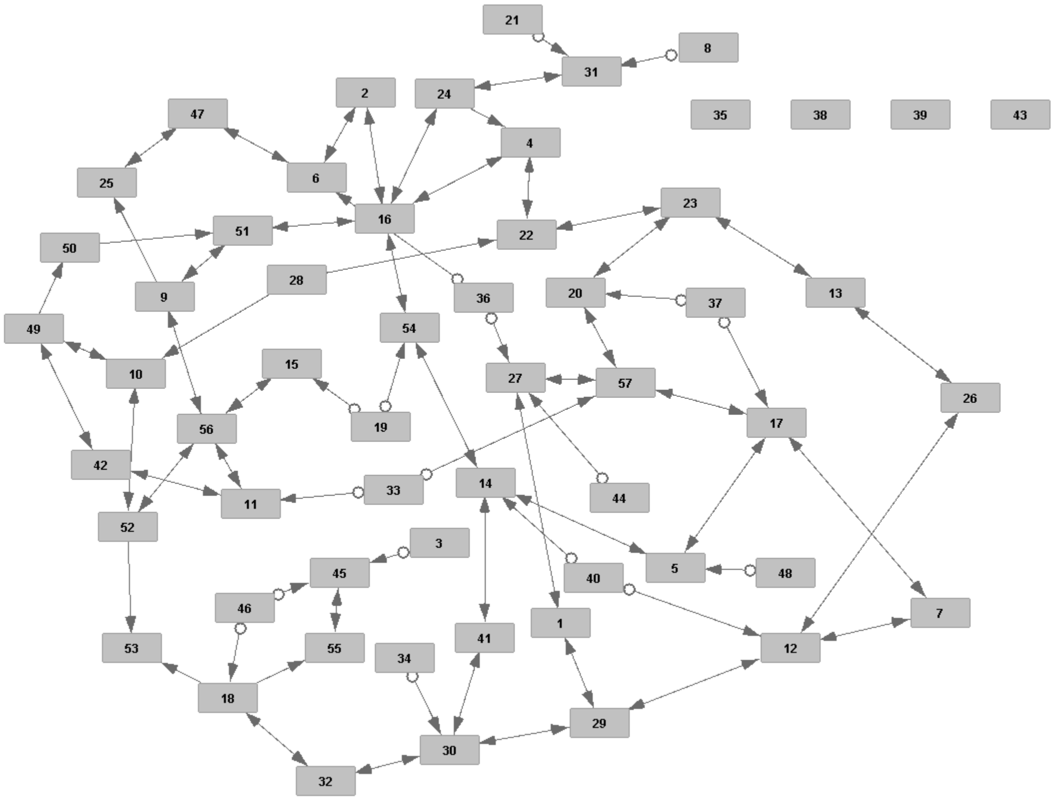


Figure 5.5: Output of Fast Causal Inference algorithm. Each node represents a taxon. Table 5.4 gives a key to plant, collembolan, and mesostigmatid taxa; fungal OTUs are T-RFLP alleles. Bidirected links indicate an unmeasured common cause of two variables, while partly directed links (those with a circle on one end, A○→B) indicate that the algorithm could not completely orient the link, but did find that A and B are connected and B is not a cause of A Spirtes et al. (1995). Drawn with Tetrad IV (http://www.phil.cmu.edu/~tetrad).

5.3.3 Causal Discovery

The output of the Fast Causal Inference algorithm is shown in Fig. 5.5. In this graph, 1.88% of possible directed connections actually exist.

Figure 5.6 shows the results of the comparison of the frequencies of inter- and intra-guild interactions to the values expected if all interaction types were equally probable. There is substantial overlap between the distributions and the difference between the medians is 0.258 (bootstrap 95% CI: -1.0472 to 1.0472). A bootstrap two-group ranked comparison

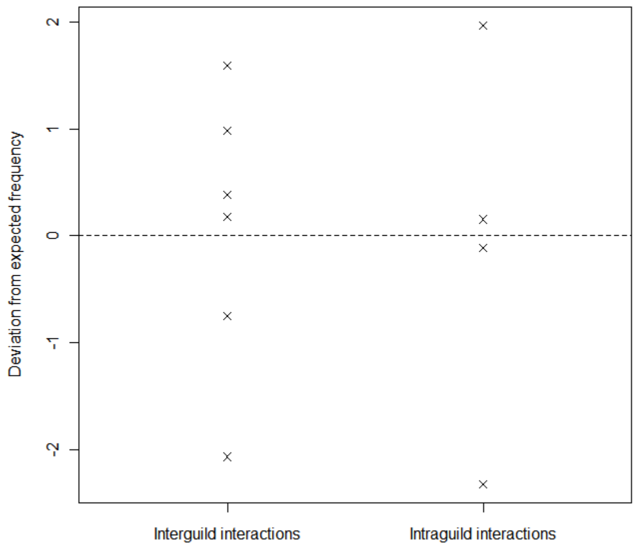


Figure 5.6: Differences between observed frequencies of inter- and intra-guild causal links and expected frequencies if all types of links were equally likely.

of the medians gave a p-value of 0.2917. While it is misleading to declare a finding of “no significant difference” based on a nonsignificant p-value in a low-power comparison such as this one (Freiman et al., 1992, Altman and Bland, 1995), in this case, there is no question of declaring a substantial difference to be statistically insignificant, as the difference between the groups is small and Figure 5.6 shows substantial overlap between the deviations from observed frequencies of inter- and intra-guild interactions. We therefore conclude that interguild interactions are approximately as common as intraguild ones, although their relative strengths remain unknown.

5.4 Discussion

5.4.1 Causal Analysis

The fact that data on abundances of various taxa along an elevation gradient produced a causal graph in which interguild interactions are approximately as common as intraguild ones links research on species distributions to that on food webs, pollination networks, and other interaction networks currently under intense study. It also supports the idea that groups of similar organisms (guilds or taxa) are not the optimal units for research in community ecology (Allen and Hoekstra, 1992, Putman, 1994). While defining the community as the biotic part of the ecosystem may be helpful, we freely admit to not having a ready answer to the question of how to best choose units for study, although some potential guidelines are discussed below. Still, we believe the question itself to be an important one for ecologists to consider.

As indicated by the frequency of bidirected and partially directed links in Figure 5.5, we found many cases in which unmeasured variables may be affecting the abundances of various taxa. While this is unsurprising given the complexity of ecological systems, it serves to identify specific topics for future research and data integration. For example, in causal analysis, the directness or indirectness of a relationship can only be discussed relative to a particular set of variables. If A affects B and B affects C but we only measure A and C, any relationship revealed would be interpreted as direct. However, it becomes indirect once B is measured and shown to be intermediate in the causal chain. This means that a comprehensive causal graph of a community can start with one or a few taxonomic groups. As more taxa and other variables are studied, their abundances can be included in the dataset subjected to causal analysis, providing mechanisms and helping to orient links. This progressive approach will gradually clarify direct and indirect relationships and can create a framework for data synthesis and Institutionalized Model-Making (Sage et al., 2003).

In future work, it will also be important to measure the effects of these interactions and examine the distribution of strong versus weak and positive versus negative causal links. The use of discrete measures of abundance will allow causal discovery algorithms that do not assume an acyclic causal structure to be applied more easily(Shipley, 2002, Pearl, 2009). Studies comparing levels of integration of different community types will also be of interest, as will the application of social network and other structural measures to causal graphs.

5.4.2 Elements of Metacommunity Structure

Most taxonomic groups, including plants, exhibited a Clementsian or quasi-Clementsian community structure, with substantial boundary conjunction. For all single guilds except fungi analyzed without Site C, Morisita's Index showed a higher degree of boundary conjunction than that expected if taxa were distributed independently of each other. In addition, the only guild pairs that exhibited higher than expected boundary conjunction were plants-fungi, fungi-collembola, and collembola-mesostigmatids. These pairs are adjacent in the food chain consisting of the four guilds studied and are thus expected to directly interact. These findings, particularly the fact that all and only pairs of directly interacting guilds had higher-than-expected amounts of boundary conjunction, support the idea that biotic interactions drive the distributions of many taxa at our site, although soil chemistry was not taken into account in the analysis.

While all triplets except fungi, collembola and mesostigmatids were Clementsian, only the plants, collembola and fungi triplet had an MI value that was higher than the weighted average of its component guilds' MI values. This is consistent with our finding of higher-than-expected MI values in trophically interacting pairs and is expected given the importance of microbes in pre-processing low-quality plant food for other consumers. On the other hand, the much lower-than-expected MI value for fungi, collembola and mesostigmatids is an exception to this trend. It may be partly related to the weak coherence exhibited by collembola (Table 5.2); the interactions among these organisms and their implications for taxon distributions should be studied more in the future. The full subcommunity consisting of all four guilds was Clementsian and had a higher-than-expected MI value, in conformity with the overall pattern.

Higher-resolution sampling may show less boundary conjunction, as Morisita's index only measures boundary conjunction and does not take into account how close non-coinciding boundaries are to each other. For high-resolution studies, it may be beneficial to use a conjunction index that takes this distance into account. Also, the size of study units must be related to the scale at which the organisms being studied live. This study did not use highly mobile organisms such as birds or flying insects because these organisms may be found in a plot without interacting with anything there. If a study is to have a chance of detecting the influence of biotic interactions on the distribution of such organisms, the sampling units must be large enough to encompass a substantial part of a typical individual's home range.

The prevalence of Clementsian and quasi-Clementsian community structure in this study is consistent with prior work using EMS analysis (Leibold and Mikkelson, 2002, Presley et al., 2009, 2010b, Presley and Willig, 2010). Indeed, Leibold and Mikkelson (2002) found that eight out of 19 datasets from Whittaker's gradient studies, including ones from the Great Smoky Mountains, support a Clementsian rather than a Gleasonian organization, having statistically significant boundary conjunction.

Since our study site is geographically close to Whittaker's (Whittaker, 1956), it is interesting that our results show predominantly Clementsian and quasi-Clementsian organization, while many of his are Gleasonian. This could be a result of differences between sites; however, other possibilities exist. Whittaker's data tables lump all sites from each elevation, which may affect the outcome of EMS analysis. Also, Whittaker's studies took place shortly after American chestnut was functionally eliminated from southern Appalachian forests, resulting in substantial changes in the abundances of other tree species (Nelson, 1955). It is possible that forest response to this fairly recent loss of a dominant species affected the patterns of co-occurrence found in his study.

Currently, individualistic community conceptions are dominant among ecologists (Putman, 1994, Lortie et al., 2004, Ricklefs, 2008). Yet studies since at least Pielou and Routledge (1976) have found evidence of boundary conjunction along environmental gradients. Others have documented both facilitation and competitive interactions (Lortie et al., 2004). As mentioned previously, studies using the Elements of Metacommunity Structure framework have frequently found Clementsian patterns with coherent species ranges and statistically significant boundary conjunction (Leibold and Mikkelson, 2002, Presley et al., 2009, 2010b, Presley and Willig, 2010).

5.5 Conclusion

We found that interguild interactions are approximately as common as intraguild ones, although their relative strengths remain unknown. Furthermore, Clementsian communities are common and sets of trophically interacting guilds have a strong tendency to show elevated levels of boundary conjunction, unlike sets of guilds that are not expected to directly interact. Together, these findings argue for the importance of biotic interactions in structuring ecological communities. Trophic interactions and plant-fungus symbioses (mutualistic and/or pathogen-host) appear particularly important but cannot be detected by research that focuses on single taxa or guilds. Future research should therefore be integrative, including several guilds.

Table 5.4. Key to Taxa in Figure 5.5

Number	Guild	Taxon
1	Plants	*Clintonia umbellulata*
2	Plants	*Hexastylis* sp.

3	Plants	*Ilex montana*
4	Plants	*Lindera benzoin*
5	Plants	*Rubus allegheneiensis*
6	Plants	*Rubus sp.*
7	Plants	*Castanea dentata*
8	Plants	*Carpinus caroliniana*
9	Plants	*Tilia americana*
10	Plants	*Betula lenta*
11	Plants	*Prunus serotina*
12	Plants	*Nyssa sylvatica*
13	Plants	*Robinia pseudoacacia*
14	Plants	*Quercus prinus*
15	Plants	*Magnolia acuminata*
16	Plants	*Cornus florida*
17	Plants	*Fraxinus pennsylvanica*
18	Plants	*Tsuga canadensis*
19	Plants	*Kalmia latifolia*
20	Plants	*Quercus rubra*
21	Plants	*Carya glabra*
22	Plants	*Carya ovalis*
23	Plants	*Acer rubrum*
24	Plants	*Rhododendron maximum*
25	Plants	*Amelanchier arborea*
26	Plants	*Oxydendrum arboreum*

27	Plants	*Acer pensylvanicum*
28	Plants	*Acer saccharinum*
29	Plants	*Liriodendron tulipifera*
30	Plants	*Quercus alba*
31	Plants	*Pinus strobus*
32	Plants	*Betula alleghaniensis*
33	Collembola	Entomobryidae
34	Collembola	Isotomidae
35	Collembola	Onychiuridae
36	Collembola	Hypogastruridae
37	Mesostigmatids	Pachylaelapidae
38	Mesostigmatids	Veigaiidae
39	Mesostigmatids	Laelapidae
40	Mesostigmatids	Ologamasidae
41	Mesostigmatids	Zerconidae
42	Mesostigmatids	Dermanyssiae
43	Mesostigmatids	Parholaspididae
44	Mesostigmatids	Rhodacaridae
45	Mesostigmatids	Parasitidae
46	Mesostigmatids	Ascidae

Chapter 6

Conclusions

> “The thing the ecologically illiterate don’t realize about an ecosystem is that it’s a system. A system! A system maintains a certain fluid stability that can be destroyed by a misstep in just one niche. A system has order, a flowing from point to point. If something dams the flow, order collapses. The untrained miss the collapse until too late. That’s why the highest function of ecology is the understanding of consequences.” –Frank Herbert, *Dune*, 1965
>
> “The thing the ecologically illiterate don’t realize about an ecosystem is that it’s a system. A system! A system maintains a certain fluid stability that is hard to destroy by a misstep in any one part. A system has order, a flowing from state to state. If something dams the flow and the environment changes, the system can’t adapt and order collapses. The untrained miss the collapse until too late. That’s why the highest function of ecology is the understanding of processes.” –Jane Shevtsov, 2006

6.1 Importance of networks and systems thinking

Any set of entities connected by links can be thought of as a network. In ecology, the links represent binary relationships between taxa or ecosystem compartments. The relationships thus represented include predation, pollination, seed dispersal, chemical transformation, or any of a number of other things. The research described in this dissertation has all been based, explicitly or implicitly, on the network concept and highlights the importance of whole-system approaches in ecology. The potential importance of indirect energy flows in model food webs and the fact that functional diversity increases the coherence of forest soil communities both lead us to conclude that ecosystems must be considered as wholes.

Such views have been expressed for as long as ecology has been a science. Forbes (1887) referred to "the impossibility of studying completely any form [taxon] out of relation to the other forms". Over a hundred years later, Yodzis (1998) demonstrated the impossibility of predicting whether a seal cull would have a positive or negative effect on a particular prey species without taking the whole food web into account.

The network approach, implicit in Forbes and explicit in Yodzis, allows us to use local information about interactions to build up a structure from which global interactions can be inferred. The network thus constructed can be studied as a whole, as in Chapter 4. There are intriguing hints, particularly from the social sciences (e.g. Granovetter, 1973, Rothenberg et al., 1998), that network structure can give us powerful explanations of natural phenomena, a possibility that should be pursued by empirical and theoretical ecologists.

6.2 Limitations of network approaches

While conceptualizing ecosystems as networks is an important part of systems ecology, we should not forget other tools. Odum (1983) gives a list of 33 "systems languages", ranging from differential equations to art. Some of the languages in his list, which is far from exhaustive, can be represented well by networks, but others cannot. In particular, while ordinary differential equation (ODE) models are often used together with networks, ODEs are far more general.

Unlike networks, differential equations can represent interactions involving more than two species. Such interactions, termed *higher-order interactions* (Wootton, 1994), are quite common. A simple example would be a situation in which a predator's success rate while hunting depends on the amount of vegetation cover available to the prey. Since such interactions are not binary, they cannot be represented as networks. However, ODEs can include higher-order interactions by incorporating all relevant species in the functions representing reproduction, growth, predation and metabolism (Kéfi et al., 2012). The effects of ecosystem engineers, organisms that extensively modify their environment in ways that change its suitability for other organisms (Jones et al., 1994), are also difficult to incorporate into networks, particularly those representing transfers of a conserved currency. Similarly, the fact that real ecosystems cover a spatial area or volume is lost in network and ODE representations but can be modeled with partial differential equations and related discrete approaches.

6.3 Community concepts and system states

Currently, individualistic community conceptions are dominant among ecologists (Putman, 1994, Lortie et al., 2004, Ricklefs, 2008). Yet studies since at least Pielou and Routledge (1976) have found evidence of boundary conjunction along environmental gradients. Others have documented both facilitation and competitive interactions (Lortie et al., 2004). As mentioned previously, studies using the Elements of Metacommunity Structure framework have frequently found Clementsian patterns with coherent species ranges and statistically significant boundary conjunction (Leibold and Mikkelson, 2002, Presley et al., 2009, 2010b, Presley and Willig, 2010). What, then, accounts for the prevalence of individualistic community concepts?

6.3.1 Metaphors for ecological communities

The Clements-Gleason debate is three quarters of a century old. Ecologists' conceptions of community organization have changed (Allen and Hoekstra, 1992, Leibold and Mikkelson, 2002, Lortie et al., 2004), yet the old debate's terms still structure our discussions of the topic.

This may be partly due to the vividness of these terms – the community as superorganism (an analogy that Clements may have used because the term “complex self-organizing system” did not exist during his time (Hagen, 1992)) versus “every species of plant is a law unto itself” (Gleason, 1926).

Metaphors and analogies can be critical to scientific understanding (Lightman, 1989). While the superorganism metaphor is the best-known one in community ecology, others have been put forth. An endnote in Allen and Hoekstra (1992) cites the early ecologist John Curtis as saying that ecological communities function as if the species are “connected by elastic bands” whose tension is “constantly shifting because, as one member of the community changes, its relationship to immediate others changes and the tension passes through the web of rubber bands”. Indeed, such a conception of communities is employed when they are modeled as systems of coupled oscillators (McCann et al., 1998, Huxel and McCann, 1998, Vandermeer, 2006, Hastings, 2010).

A better metaphor for communities may be that suggested by the word “community” itself and by the cognateness of the words “ecology” and “economy”. Consider an urban neighborhood whose residents interact in certain economic and social contexts. Like most neighborhoods, it has fuzzy spatial boundaries and interactions with the larger city are very common. It is constantly changing but retains an integrity through most – but not all – possible changes. Many people living in the neighborhood don’t directly interact with each other, but the interactions that do take place may be critical. Still, residents are changed to some extent by living there and their interactions are modified by the social context.

The analogy presented here is far from perfect. (In particular, since the physical structure of terrestrial communities is made up largely of living plants, the analogy would work better if the neighborhood’s buildings could not be reoccupied upon the death of the owner.) However, it may still prove fruitful as a source of ideas and pedagogical tool, as well as a way to connect ecological communities with other complex systems.

6.3.2 What is the system state?

Ricklefs (2008) wrote one of the sharpest critiques in recent years of non-individualistic community concepts, explicitly describing the ecological community as “an epiphenomenon that has little explanatory power”. The research described in Chapter 5, as well as that of many others (e.g. Pielou and Routledge, 1976, Leibold and Mikkelson, 2002), provides evidence for interactive communities. However, as long as ecologists define the state of a community or ecosystem as the vector of abundances of its component species or compartments, we will be vulnerable to the charge that the system we are studying is merely an epiphenomenon of population distributions. We need to find properties of ecological systems that will allow us to define macrostates of entire communities and ecosystems that are more than lists of the states of their parts.

If such properties are chosen well, whole systems can be studied without detailed knowledge of their parts. Indeed, such knowledge may be a distraction. The mathematician René Thom wrote, “A knowledge of the fine structure, molecules for a fluid, cells for an animal, is practically irrelevant for understanding the global structure... of the total system. For instance, the final structure of a theory like Fluid Mechanics does not depend on whether one takes as the basic concept molecules or a continuous fluid.” (René Thom, “Structuralism and Biology”, quoted in Garfinkel (1981)) This would not be possible if the state of the fluid was taken to be the set of positions and velocities of all its molecules.

Ecologists have started looking for these kinds of properties in research on indicators of ecosystem health. For example, Ulanowicz (1997) has proposed that ascendency be used to determine if an ecosystem is under stress. However, ascendency is a network property and fairly detailed knowledge of ecosystem structure and function is required to compute it. Jørgensen and Nielsen (2007) promote the use of eco-exergy, a combination of the thermodynamic distance of ecosystem components from chemical equilibrium and the information stored in the DNA of the organisms in the ecosystem. This measure, being a sum over components, shares the inadequacies of a vector of states and, unlike ascendency, does not incorporate information about ecosystem organization. Schneider and Kay (1994) argue that a terrestrial ecosystem's temperature should be used as an index of its development, but this is almost completely determined by plant evapotranspiration, so it is hard to see it as an adequate ecosystem descriptor. However, the idea of focusing on ecosystem outputs may well be a productive one.

As the term "ecosystem health" implies, research in this area has a strongly applied flavor, despite the large amounts of theory it uses. Holistically-oriented community and ecosystem ecologists with more basic interests could do worse than to spend some time figuring out what exactly they want to know about their study systems and how system macrostates might be described.

The network approach to communities and ecosystems provides ecologists with a powerful set of tools and ideas. Together with other languages and a focus on whole systems, it may enable us to develop a better understanding of natural processes and consequences.

Appendices

Appendix A

Matlab Code for Dynamic Environ Approximation

```
function indirect_mov(TimeCourseIn,Mat_input,Mat_type,WindowSize,Dt)
% This function estimates an indirect flow fraction for almost any
% ecological network.
%   Before running the function, use an ODE solver to create the x
% matrix containing time series for each variable. Enter the name
% of this matrix for "TimeCourseIn".
%   If the stock-normalized flow values do not remain constant
% (i.e., the underlying model is non-stationary or non-linear),
% create an array of these values and enter its name after the
% name of your time series. (If the system is linear and
% stationary, C is a simple matrix.) Alternatively, you may create
% an array of F matrices. Enter the name of the C or F array for
% "Mat_input" and 0 (for C) or 1 (for F) for Mat_type.
```

```
%   Choose a length ("WindowSize") for your product series. The
% length refers to the number of values in your time series that
% should be used in the product series. Finally, specify the step
% size ("Dt") you used when running the simulation in your ODE
% solver. (For EcoNet simulations, Dt=1.)
%   This function is only compatible with MATLAB 7.1 or higher.
%   Written by Jane Shevtsov and Caner Kazanci.

% Get array dimensions.
Length=size(TimeCourseIn, 1);
Width=size(Mat_input, 1);
Iter=ceil(Length*Dt)-WindowSize;
% If C or F is a 2-D matrix, convert it to a 3-D array.
if (size(Mat_input,3) == 1)
   Matlarge(Width, Width, Length) = 0;
   for i = 1:Length
    Matlarge(:,:,i) = Mat_input;
  end
else
    Matlarge=Mat_input;
end

% Delete intermediate time points.
Steps=1/Dt;
Pick=[1:Steps:Length]';
LengthRed=ceil(Length*Dt);
```

```

TimeCourse=zeros(LengthRed,Width);
Mat=zeros(Width,Width,LengthRed);
for i=1:LengthRed
    Mat(:,:,i)=Matlarge(:,:,Pick(i,1));
    TimeCourse(i,:)=TimeCourseIn(Pick(i,1) ,:);
end

% Create the G array.
if Mat_type==0  %For C matrix/array.
    Fbar=zeros(Width, Width, LengthRed);
    for k=1:LengthRed
        Fbar(:,:,k)=Mat(:,:,k)*diag(TimeCourse(k,:));
        T(k,:)=-1*diag(Fbar(:,:,k));
        F(:,:,k)=Fbar(:,:,k)+diag(T(k,:));
        G(:,:,k)=F(:,:,k)/diag(T(k,:));
    end
else  %For F array.
    for k=1:Length
    T(k,:)=-1*diag(Mat(:,:,k));
    G(:,:,k)=Mat(:,:,k)/diag(T(k,:)) + eye(Width);
    end
end

% Initialize products and sums of G.
global N
```

```
Gprod=zeros(Width, Width, Iter+1);
N=zeros(Width, Width, Iter+1);

% Compute Gprod and N.
for j = 1:(Iter+1)
  Gprod(:,:,j) = eye(Width);
  for i = 1:WindowSize
        %Multiply and add to obtain sum of matrix products.
    Gprod(:,:,j) = Gprod(:,:,j) * G(:,:,i+j-1);
    N(:,:,j) = N(:,:,j) + Gprod(:,:,j);
  end
  N(:,:,j)=N(:,:,j)+eye(Width);
end

% Compute indirect flows.
Ind=zeros(Width,Width,Iter);
for i=1:Iter
    Ind(:,:,i)=N(:,:,i)-G(:,:,i)-eye(Width);
end

% Compute indirect to total flow ratios.
CountZeros=zeros(1,Iter);
CountOnes=zeros(1,Iter);
Ratio=zeros(Width,Width,Iter);
% Nested loops allow calculation to be done 1 entry at a time,
% to avoid division by 0.
```

```
for k = 1:Iter
    for j = 1:Width
        for i = 1:Width
            if N(i,j,k)==0 %Avoiding division by zero.
                Ratio(i,j,k)=0;
                CountZeros(1,k)=CountZeros(1,k)+1;
            else  %Calculating indirect/total ratio.
                Ratio(i,j,k)=Ind(i,j,k)/N(i,j,k);
                if Ratio(i,j,k)==1
                    CountOnes(1,k)=CountOnes(1,k)+1;
                end
            end
        end
    end
end

% Create movie of I/T ratios.
M=moviein(Iter);
for i=1:Iter
    RatioFlip(:,:,i)=Ratio(Width:-1:1,:,i);
end
global RatioD
RatioD=zeros(Width+1,Width+1,Iter);
for i=1:Iter
    RatioD(:,:,i)=[RatioFlip(:,:,i) RatioFlip(:,Width,i);...
        RatioFlip(Width,:,i) RatioFlip(Width,Width,i)];
```

```
    pcolor(RatioD(:,:,i));
    colormap gray
    set(gca,'xtick',1:1:Width)
    set(gca,'ytick',1:1:Width)
    set(gca,'YDir', 'reverse')
    xlabel donor
    ylabel recipient
    colorbar
    M(:,i)=getframe;
  end
movie(M,3)

% Generate summary statistics for indirect flow fractions.
Mid=zeros(1,Iter);
Maximum=zeros(1,Iter);
Minn=zeros(1,Iter);
Mean=zeros(1,Iter);
for i=1:Iter
    Mid(1,i) = median(median(Ratio(:,:,i)));
    Maximum(1,i) = max(max(Ratio(:,:,i)));
    Minn(1,i) = min(min(Ratio(:,:,i)));
    Mean(1,i) = mean(mean(Ratio(:,:,i)));
end

% Plot summary statistics.
Time=1:Iter;
```

```
figure
plot(Time, Minn, 'kx', Time, Mid, 'k--+', ...
         Time, Mean, 'k-.', Time, Maximum, 'ks')
legend('minimum', 'median', 'mean', 'maximum')
xlabel('time')
ylabel('Indirect␣flow␣fraction')

% Display numbers of unlinked pairs and pairs linked only
% by indirect flows.
Ionly=CountOnes(1,Iter-1); Uncon=CountZeros(1,Iter-1);
[num2str(Ionly), '␣pairs␣of␣compartments␣are␣linked␣only␣by␣
    indirect␣paths.']
[num2str(Uncon), '␣pairs␣of␣compartments␣are␣unconnected.']
```

Appendix B

Variables Used in CART Analysis

27 variables were used as potential predictors in the CART analysis. They included nominal and actual size and connectance, mean path length, graph diameter (the largest minimum distance between two nodes), and clustering coefficient (the probability that two nodes that share a neighbor are themselves connected). Structural cycling was quantified as either dominant or maximum eigenvalue. Several properties of the degree distribution of the adjacency matrix were also used as potential predictors. These were the mean, median, maximum, minimum, variance and mean absolute deviation from the median (MADAM). The same measures were applied to trophic levels. Finally, several explicitly ecological quantities were used. These were the mean generality (the number of prey a species has), mean vulnerability (number of predators), and the fraction of species at top, bottom and intermediate trophic levels.

Bibliography

Agnarsson, G. and Greenlaw, R. 2006. Graph theory: modeling, applications, and algorithms. Prentice Hall.

Agnew, J. A. 2002. Making political geography. Oxford University Press, New York.

Allen, T. F. H. and Hoekstra, T. W. 1992. Toward a unified ecology. Columbia University Press, New York.

Allesina, S., Alonso, D., and Pascual, M. 2008. A general model for food web structure. *Science* 320:658–661.

Allesina, S. and Tang, S. 2012. Stability criteria for complex ecosystems. *Nature* 483:205–208.

Altman, D. G. and Bland, J. M. 1995. Absence of evidence is not evidence of absence. *British Medical Journal* 311:485.

Anderson, W. N. and Morley, T. D. 1971. Eigenvalues of the Laplacian of a graph. Technical report, NASA.

Austin, M. 1985. Continuum concept, ordination methods, and niche theory. *Annual Review of Ecology and Systematics* 16:39–61.

Banasek-Richter, C., Bersier, L.-F., Cattin, M.-F., Baltensperger, R., Gabriel, J.-P., Merz, Y., Ulanowicz, R. E., Tavares, A. F., Williams, D. D., de Ruiter, P. C., Winemiller, K. O., and Naisbit, R. E. 2009. Complexity in quantitative food webs. *Ecology* 90:1470–7.

Barabási, A.-L. and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Bascompte, J., Jordano, P., Olesen, J. M., and Melián, C. 2003. The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences* 100:9383–9387.

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

Berlow, E. L., Dunne, J. A., Martinez, N. D., Stark, P. B., Williams, R. J., and Brose, U. 2009. Simple prediction of interaction strengths in complex food webs. *Proceedings of the National Academy of Sciences* 106:187–91.

Bersier, L.-F. 2007. A history of the study of ecological networks, pp. 365–421. *In* F. Kepes (ed.), Biological Networks, chapter 11. World Scientific Publishing Company, Hackensack, NJ.

Boring, L. R., Monk, C. D., and Swank, W. T. 1981. Early regeneration of a clear-cut southern Appalchian forest. *Ecology* 62:1244–1253.

Borrett, S. R. and Freeze, M. 2011. Reconnecting environs to their environment. *Ecological Modelling* 222:2393–2403.

Borrett, S. R. and Salas, A. 2010. Evidence for resource homogenization in 50 trophic ecosystem networks. *Ecological Modelling* 221:1710–1716.

Borrett, S. R., Whipple, S. J., and Patten, B. C. 2010. Rapid development of indirect effects in ecological networks. *Oikos* 119:1136–1148.

Borrett, S. R., Whipple, S. J., Patten, B. C., and Christian, R. 2006. Indirect effects and distributed control in ecosystems: Temporal variation of indirect effects in a seven-compartment model of nitrogen flow in the Neuse River Estuary, USA – Time series analysis. *Ecological Modelling* 194:178–188.

Bossenbroek, J. M., Wagner, H. H., and Wiens, J. A. 2005. Taxon-dependent scaling: beetles, birds, and vegetation at four North American grassland sites. *Landscape Ecology* 20:675–688.

Briand, F. and Cohen, J. E. 1984. Community food webs have scale-invariant structure. *Nature* 307:264–267.

Brose, U., Jonsson, T., Berlow, E. L., Warren, P., Banasek-Richter, C., Bersier, L.-F., Blanchard, J. L., Brey, T., Carpenter, S. R., Blandenier, M., and Others 2006. Consumer-resource body-size relationships in natural food webs. *Ecology* 87:2411–2417.

Brown, J. H. 1999. Macroecology: progress and prospect. *Oikos* 87:3–14.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and West, G. B. 2004. Toward a metabolic theory of ecology. *Ecology* 85:1771–1789.

Bruno, J. F., Fridley, J., Bromberg, K., and Bertness, M. 2005. Insights into biotic interactions from studies of species invasions, pp. 13–40. *In* D. F. Sax, J. J. Stachowicz, and S. D. Gaines (eds.), Species invasions: insights into ecology, evolution, and biogeography, chapter 1. Sinauer Associates, Sunderland, MA.

Cattin, M.-F., Bersier, L.-F., Banasek-Richter, C., Baltensperger, R., and Gabriel, J.-P. 2004. Phylogenetic constraints and adaptation explain food-web structure. *Nature* 427:835–839.

Christian, R. and Luczkovich, J. J. 1999. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecological Modelling* 117:99–124.

Clements, F. 1936. Nature and structure of the climax. *Journal of Ecology* 24:252–284.

Cohen, J. E. 1977. Ratio of prey to predators in community food webs. *Nature* 270:165–167.

Cohen, J. E. 1978. Food webs and niche space. Princeton University Press, Princeton, NJ.

Cohen, J. E. and Briand, F. 1984. Trophic links of community food webs. *Proceedings of the National Academy of Sciences* 81:4105–9.

Cohen, J. E., Briand, F., and Newman, C. M. 1990. Community food webs: data and theory. Springer-Verlag, New York.

Cohen, J. E. and Newman, C. 1985. A stochastic theory of community food webs: I. Models and aggregated data. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 224:421–448.

Coleman, D. C., Crossley Jr, D. A., and Hendrix, P. F. 2004. Fundamentals of Soil Ecology. Academic Press, Burlington, MA, 2nd edition.

Cousins, S. 1987. The decline of the trophic level concept. *Trends in Ecology and Evolution* 2:312–6.

Cross, W. F., Wallace, J. B., and Rosemond, A. D. 2007. Nutrient enrichment reduces constraints on material flows in a detritus-based food web. *Ecology* 88:2563–75.

Csermely, P. 2006. Weak links: stabilizers of complex systems from proteins to social networks. Springer, New York.

Dame, R. F. and Patten, B. C. 1981. Analysis of energy flows in an intertidal oyster reef. *Marine Ecology Progress Series* 5:115–124.

Dickie, I. A. and FitzJohn, R. G. 2007. Using terminal restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi: a methods review. *Mycorrhiza* 17:259–270.

Douglass, J. E. and Swank, W. T. 1975. Effects of management practices on water quality and quantity: Coweeta Hydrological Laboratory, North Carolina. Technical report.

Dunne, J. A. 2006. The network structure of food webs, pp. 27–86. *In* M. Pascual and J. A. Dunne (eds.), Ecological Networks: Linking Structure to Dynamics in Food Webs, chapter 2. Oxford University Press, USA.

Dunne, J. A., Williams, R. J., and Martinez, N. D. 2002. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters* 5:558–567.

Eagle, N., Macy, M., and Claxton, R. 2010. Network diversity and economic development. *Science* 328:1029–31.

Edwards, I. P., Cripliver, J. L., Gillespie, A. R., Johnsen, K. H., Scholler, M., and Turco, R. F. 2004. Nitrogen availability alters macrofungal basidiomycete community structure in optimally fertilized loblolly pine forests. *New Phytologist* 162:755–770.

Fath, B. D. 2004. Network analysis applied to large-scale cyber-ecosystems. *Ecological Modelling* 171:329–337.

FATH, B. D. 2007. Network mutualism: Positive community-level relations in ecosystems. *Ecological Modelling* 208:56–67.

FATH, B. D. AND PATTEN, B. C. 1998. Network synergism: emergence of positive relations in ecological systems. *Ecological Modelling* 107:127–143.

FATH, B. D. AND PATTEN, B. C. 1999a. Quantifying resource homogenization using network flow analysis. *Ecological Modelling* 123:193–205.

FATH, B. D. AND PATTEN, B. C. 1999b. Review of the foundations of network environ analysis. *Ecosystems* 2:167–179.

FATH, B. D., PATTEN, B. C., AND CHOI, J. S. 2001. Complementarity of ecological goal functions. *Journal of Theoretical Biology* 208:493–506.

FORBES, S. A. 1880. The food of fishes (Acanthopteri). *Bulletin of the Illinois Laboratory of Natural History* 1:18–65.

FORBES, S. A. 1887. The lake as a microcosm. *Bulletin of the Peoria Scientific Association* pp. 77–87.

FOX, J. W. 2006. Current food web models cannot explain the overall topological structure of observed food webs. *Oikos* 115:97–109.

FREIMAN, J. A., CHALMERS, T. C., SMITH, H., AND KUEBLER, R. R. 1992. The importance of beta, the Type II error, and sample size in the design and interpretation of the randomized controlled trial: Survey of two sets of "negative" trials, p. 449. *In* J. C. Bailar and F. Mosteller (eds.), Medical uses of statistics, chapter 19. NEJM Books, 2nd edition.

GARDES, M. AND BRUNS, T. D. 1993. ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2:113–118.

Garfinkel, A. 1981. Forms of explanation: rethinking the questions in social theory. Yale University Press, New Haven Conn.

Gleason, H. A. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53:7–26.

Goldwasser, L. and Roughgarden, J. 1993. Construction and analysis of a large Caribbean food web. *Ecology* 74:1216–1233.

Good, P. I. 2005. Introduction to Statistics Through Resampling Methods and R/S-PLUS. Wiley-Interscience.

Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology* 78:1360–1380.

Gurney, W. S. C. and Nisbet, R. M. 1998. Ecological Dynamics. Oxford University Press, USA.

Hagen, J. B. 1992. An entangled bank : the origins of ecosystem ecology. Rutgers University Press, New Brunswick, NJ.

Hallam, T. and Antonios, M. 1985. An environ analysis for nonlinear compartment models. *Bulletin of Mathematical Biology* 47:739–748.

Hastings, A. 2010. Timescales, dynamics, and ecological understanding. *Ecology* 91:3471–3480.

Havens, K. 1992. Scale and structure in natural food webs. *Science* 257:1107–1109.

Heino, J. 2005. Metacommunity patterns of highly diverse stream midges: gradients, chequerboards, and nestedness, or is there only randomness? *Ecological Entomology* 76:417–599.

Heino, J. 2009. Species co-occurrence, nestedness and guild-environment relationships in stream macroinvertebrates. *Freshwater Biology* 54:1947–1959.

Herbert, F. 1965. Dune. Chilton Books, Philadelphia.

Higashi, M., Burns, T. P., and Patten, B. C. 1993a. Network trophic dynamics: the tempo of energy movement and availability in ecosystems. *Ecological Modelling* 66:43–64.

Higashi, M. and Patten, B. C. 1986. Further aspects of the analysis of indirect effects in ecosystems. *Ecological modelling* 31:69–77.

Higashi, M. and Patten, B. C. 1989. Dominance of indirect causality in ecosystems. *American Naturalist* 133:288–302.

Higashi, M., Patten, B. C., and Burns, T. P. 1991. Network trophic dynamics: an emerging paradigm in ecosystems ecology, pp. 117–154. *In* M. Higashi and T. P. Burns (eds.), Theoretical Studies of Ecosystems: The Network Perspective‘, chapter 5. Cambridge University Press, New York.

Higashi, M., Patten, B. C., and Burns, T. P. 1993b. Network trophic dynamics: the modes of energy utilization in ecosystems. *Ecological Modelling* 66:1–42.

Hippe, P. 1983. Environ analysis of linear compartmental systems: the dynamic, time-invariant case. *Ecological Modelling* 19:1–26.

Hoagland, B. and Collins, S. L. 1997. Gradient models, gradient analysis, and hierarchical structure in plant communities. *Oikos* 78:23–30.

Hubbell, S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.

Hunt, C. B. 1973. Natural Regions of the United States and Canada. W.H.Freeman & Co Ltd.

Huxel, G. R. and McCann, K. S. 1998. Food web stability: the influence of trophic flows across habitats. *American Scientist* 152:460–9.

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L. E., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., Lauridsen, R. B., Ledger, M. E., Lewis, H. M., Olesen, J. M., van Veen, F. J. F., Warren, P. H., and Woodward, G. 2009. Ecological networks – beyond food webs. *Journal of Animal Ecology* 78:253–269.

Johnson, T. 1985. A continuous Leontief dynamic input-output model. *Papers in Regional Science* 56:177–188.

Jones, C. G., Lawton, J. H., and Shachak, M. 1994. Organisms as ecosystem engineers. *Oikos* 69:373–386.

Jørgensen, S. E., Fath, B. D., Bastianoni, S., Marques, J. C., Müller, F., Nielsen, S. N., Patten, B. C., Tiezzi, E., and Ulanowicz, R. E. 2007. A new ecology : systems perspective. Elsevier, Boston, MA.

Jørgensen, S. E. and Nielsen, S. N. 2007. Application of exergy as thermodynamic indicator in ecology. *Energy* 32:673–685.

Justus, J. 2008. Complexity, diversity, and stability, pp. 321–350. *In* S. Sarkar and A. Plutynsky (eds.), A Companion to the Philosophy of Biology, chapter 18. Blackwell Publishing.

Kalisch, M., Maechler, M., and Colombo, D. 2011. pcalg: Estimation of CPDAG/-PAG and causal inference using the IDA algorithm.

Kazanci, C. 2007. EcoNet: A new software for ecological modeling, simulation and network analysis. *Ecological Modelling* 208:3–8.

Kéfi, S., Berlow, E. L., Wieters, E. a., Navarrete, S. a., Petchey, O. L., Wood, S. a., Boit, A., Joppa, L. N., Lafferty, K. D., Williams, R. J., Martinez, N. D., Menge, B. a., Blanchette, C. a., Iles, A. C., and Brose, U. 2012. More than a meal…integrating non-feeding interactions into food webs. *Ecology Letters* 15:291–300.

Kendrick, D. 1972. On the Leontief dynamic inverse. *Quarterly Journal of Economics* 86:693–696.

Kirsten, W. 1983. Organic Elemental Analysis: Ultramicro, Micro, and Trace Methods. Academic Press/Harcourt Brace Jovanovich, New York.

Krapfl, K. J., Holzmueller, E. J., and Jenkins, M. A. 2011. Early impacts of hemlock woolly adelgid in Tsuga canadensis forest communities of the southern Appalachian Mountains. *Journal of the Torrey Botanical Society* 138:93–106.

Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E., and Taylor, W. W. 2003. Compartments revealed in food-web structure. *Nature* 426:282–5.

Krebs, C. J. 2001. Ecology: the experimental analysis of distribution and abundance. Benjamin Cummings, 5th edition.

Landeweert, R., Leeflang, P., Smit, E., and Kuyper, T. W. 2005. Diversity of an ectomycorrhizal fungal community studied by a root tip and total soil DNA approach. *Mycorrhiza* 15:1–6.

Lane, N. 2009. Life ascending: the ten great inventions of evolution. W.W. Norton, New York.

Leibold, M. A. and Mikkelson, G. M. 2002. Coherence, species turnover, and boundary clumping: elements of meta-community structure. *Oikos* 23:237–250.

Leontief, W. W. 1936. Quantitative input and output relations in the economic systems of the United States. *The Review of Economics and Statistics* 18:105–125.

Leontief, W. W. 1966. Input-output economics. Oxford University Press, New York.

Leontief, W. W. 1970. The dynamic inverse. *In* Proceedings of the Fourth International Conference on Input-Output Techniques, Geneva, 8-12 January 1968: Published in Honor of Wassily Leontief.

Levine, S. 1980. Several measures of trophic structure applicable to complex food webs. *Journal of Theoretical Biology* 83:195–207.

Lightman, A. P. 1989. Magic on the mind: Physicists' use of metaphor. *The American Scholar* 58:97–101.

Lindeman, R. L. 1942. The trophic-dynamic aspect of ecology. *Ecology* 23:399–417.

Lortie, C. J., Brooker, R. W., Choler, P., and Kikvidze, Z. 2004. Rethinking plant community theory. *Oikos* 107:433–438.

MacArthur, R. 1955. Fluctuations of animal populations and a measure of community stability. *Ecology* 36:533–536.

Martinez, N. D. 1991. Artifacts or attributes? Effects of resolution on the Little Rock Late food web. *Ecological Monographs* 61:367–392.

Martinez, N. D. 1992. Constant connectance in community food webs. *American Naturalist* 140:1208–1218.

Martinez, N. D. 1993. Effect of scale on food web structure. *Science* 260:242–243.

Martinez, N. D. and Cushing, L. J. 2005. Additional model complexity reduces fit to complex food-web structure, pp. 87–89. *In* M. Pascual and J. A. Dunne (eds.), Ecological networks: linking structure to dynamics in food webs, chapter Box A. Oxford University Press, New York.

Matis, J. and Patten, B. C. 1981. Environ analysis of linear compartmental systems: the static, time invariant case. *Bulletin of the International Statistical Institute* 48:527–565.

Matis, J. H., Patten, B. C., and White, G. C. 1979. Compartmental analysis of ecosystem models. International Cooperative Publishing House, Fairland Md.

May, R. M. 1972. Will a large complex system be stable? *Nature* 238:413–4.

May, R. M. 1974. Stability and Complexity in Model Ecosystems. Princeton University Press.

McArdle, B. 1988. The structural relationship: regression in biology. *Canadian Journal of Zoology* 66:2329–2339.

McArdle, B. 2003. Lines, models, and errors: regression in the field. *Limnology and Oceanography* 48:1363–1366.

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

McRae, B. H. and Beier, P. 2007. Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences* 104:19885–90.

Miehls, A. L. J., Mason, D. M., Frank, K. a., Krause, A. E., Peacor, S. D., and Taylor, W. W. 2009. Invasive species impacts on ecosystem structure and function: A comparison of the Bay of Quinte, Canada, and Oneida Lake, USA, before and after zebra mussel invasion. *Ecological Modelling* 220:3182–3193.

Mohar, B. 1991. The Laplacian spectrum of graphs, pp. 871–898. *In* Y. Alavi, G. Chartrand, O. Oellermann, and A. Schwenk (eds.), Graph theory, combinatorics, and applications, volume 2. Wiley.

Montoya, J. M. and Solé, R. V. 2002. Small world patterns in food webs. *Journal of Theoretical Biology* 214:405–412.

Nelson, N. 1987. An acid-persulfate digestion procedure for determination of phosphorus in sediments. *Communications in Soil Science & Plant Analysis* 18:359–369.

Nelson, T. C. 1955. Chestnut replacement in the Southern highlands. *Ecology* 36:352.

Neutel, A.-M., de Ruiter, P. C., and Heesterbeek, J. 2002. Stability in real food webs: weak links in long loops. *Science* 296:1120–1123.

Newman, E. 1988. Mycorrhizal links between plants: their functioning and ecological significance. *Advances in ecological research* 18:243–270.

Newman, M. E. J. 2011. Communities, modules and large-scale structure in networks. *Nature Physics* 8:25–31.

Odum, E. P. 1968. Energy flow in ecosystems: a historical review. *American Zoologist* 8:11–18.

Odum, E. P. 1969. The strategy of ecosystem development. *Science* 164:262–270.

Odum, H. T. 1957. Trophic structure and productivity of Silver Springs, Florida. *Ecological Monographs* 27:55–112.

Odum, H. T. 1960. Ecological potential and analogue circuits for the ecosystem. *American Scientist* 48:1–8.

Odum, H. T. 1971. Environment, power and society. Wiley-Interscience, New York.

Odum, H. T. 1983. Systems ecology: an introduction. Wiley, New York.

Oksanen, J. and Minchin, P. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling* 157:119–129.

Patten, B. C. 1978. Systems approach to the concept of environment. *Ohio Journal of Science* 78:206–222.

Patten, B. C. 1984. Toward a theory of the quantitative dominance of indirect effects in ecosystems. *Verhalt Gesellschaft für Ökologie* 13:271–284.

Patten, B. C. 1985. Energy cycling in the ecosystem. *Ecological Modelling* 28:1–71.

Patten, B. C. 1986. Energy cycling, length of food chains, and direct versus indirect effects in ecosystems. *Canadian Bulletin of Fisheries and Aquatic Science* 213:119–138.

Patten, B. C. 1990. Environ theory and indirect effects: a reply to Loehle. *Ecology* 71:2386–2393.

Patten, B. C. 1991. Network ecology: indirect determination of the life-environment relationship in ecosystems, pp. 288–351. *In* M. Higashi and T. P. Burns (eds.), Theoretical Studies of Ecosystems: The Network Perspective. Cambridge University Press, New York.

Patten, B. C. 1992. Energy, emergy and environs. *Ecological Modelling* 62:29–69.

Patten, B. C. 1995. Network integration of ecological extremal principles: exergy, emergy, power, ascendency, and indirect effects. *Ecological Modelling* 79:75–84.

Patten, B. C. and Auble, G. T. 1981. System theory of the ecological niche. *American Naturalist* 117:893–922.

Patten, B. C., Bosserman, R., Finn, J., and Cale, W. 1976. Propagation of cause in ecosystems. *Systems analysis and simulation in ecology* 4:457–579.

PATTEN, B. C., HIGASHI, M., AND BURNS, T. P. 1990. Trophic dynamics in ecosystem networks: Significance of cycles and storage. *Ecological Modelling* 51:1–28.

PATTEN, B. C. AND MATIS, J. 1982. The water environs of Okefenokee Swamp: An application of static linear environ analysis. *Ecological Modelling* 16:1–50.

PATTEN, B. C. AND WITKAMP, M. 1967. Systems analysis of cesium-134 kinetics in terrestrial microcosms. *Ecology* 48:813–824.

PEARL, J. 2009. Causality: Models, Reasoning and Inference. Cambridge University Press, New York, 2nd edition.

PEECH, M. 1965. Hydrogen-ion activity, pp. 914–925. *In* C. Black (ed.), Methods of Soil Analysis, Part 2: Chemical and Microbiological Properties. American Soc. Agron., Madison, WI.

PETCHEY, O. L., BECKERMAN, A. P., RIEDE, J. O., AND WARREN, P. H. 2008. Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences* 105:4191–6.

PFEFFER, P. E., DOUDS JR, D., SCHWARTZ, D., AND SHACHAR-HILL, Y. 2004. The fungus does not transfer carbon to or between roots in an arbuscular mycorrhizal symbiosis. *New Phytologist* 163:617–627.

PIELOU, E. AND ROUTLEDGE, R. 1976. Salt marsh vegetation: latitudinal gradients in the zonation patterns. *Oecologia* 24:311–321.

PIMM, S. L. AND LAWTON, J. 1980. Are food webs divided into compartments? *Journal of Animal Ecology* 49:879–898.

PIMM, S. L., LAWTON, J., AND COHEN, J. E. 1991. Food web patterns and their consequences. *Nature* 350:669–674.

Pinnegar, J. K., Blanchard, J. L., Mackinson, S., Scott, R., and Duplisea, D. 2005. Aggregation and removal of weak-links in food-web models: system stability and recovery from disturbance. *Ecological Modelling* 184:229–248.

Pires, M. M., Prado, P. I., and Guimarães, P. R. 2011. Do food web models reproduce the structure of mutualistic networks? *PLoS One* 6:e27280.

Polis, G. A. 1991. Complex trophic interactions in deserts: An empirical critique of food-web theory. *American Naturalist* 138:123–155.

Potterat, J. J., Rothenberg, R. B., and Muth, S. Q. 1999. Network structural dynamics and infectious disease propagation. *International Journal of STD & AIDS* 10:182–185.

Presley, S. J., Higgins, C. L., López-González, C., and Stevens, R. D. 2009. Elements of metacommunity structure of Paraguayan bats: multiple gradients require analysis of multiple ordination axes. *Oecologia* 160:781–93.

Presley, S. J., Higgins, C. L., and Willig, M. R. 2010a. A comprehensive framework for the evaluation of metacommunity structure. *Oikos* 119:908–917.

Presley, S. J. and Willig, M. R. 2010. Bat metacommunity structure on Caribbean islands and the role of endemics. *Global Ecology and Biogeography* 19:185–199.

Presley, S. J., Willig, M. R., Bloch, C., Castro-Arellano, I., Higgins, C. L., and Klingbeil, B. 2010b. A complex metacommunity structure for gastropods along an elevational gradient. *Biotropica* pp. 1–9.

Proulx, S., Promislow, D., and Phillips, P. 2005. Network thinking in ecology and evolution. *Trends in Ecology and Evolution* 20:345–353.

Putman, R. 1994. Community ecology. Springer.

PUTNAM, H. 1991. The 'corroboration' of theories, pp. 121–137. *In* R. Boyd, P. Gasper, and J. D. Trout (eds.), The philosophy of science, chapter 6. MIT Press.

QUINN, G. P. AND KEOUGH, M. J. 2002. Experimental design and data analysis for biologists. Cambridge University Press, New York.

R DEVELOPMENT CORE TEAM 2009. R: A language and environment for statistical computing.

RICKLEFS, R. E. 2008. Disintegration of the ecological community. *American Naturalist* 172:741–50.

ROBINSON, D. AND FITTER, A. 1999. The magnitude and control of carbon transfer between plants linked by a common mycorrhizal network. *Journal of Experimental Botany* 50:9–13.

ROSSBERG, A. G., MATSUDA, H., AMEMIYA, T., AND ITOH, K. 2005. An explanatory model for food-web structure and evolution. *Ecological Complexity* 2:312–321.

ROTHENBERG, R. B., POTTERAT, J. J., WOODHOUSE, D. E., MUTH, S. Q., DARROW, W. W., AND KLOVDAHL, A. S. 1998. Social network dynamics and HIV transmission. *AIDS* 12:1529–36.

SAGE, R. W., PATTEN, B. C., AND SALMON, P. A. 2003. Institutionalized model-making and ecosystem-based management of exploited resource populations: a comparison with instrument flight. *Ecological Modelling* 170:107–128.

SANTAMARIA, L. AND RODRIGUEZ-GIRONÉS, M. A. 2007. Linkage rules for plant-pollinator networks: trait complementarity or exploitation barriers? *PLoS Biology* 5:e31.

SCHMIDT-NIELSEN, K. 1984. Scaling: Why is Animal Size so Important? Cambridge University Press.

Schneider, E. D. and Kay, J. 1994. Life as a manifestation of the second law of thermodynamics. *Mathematical and Computer Modelling* 19:25–48.

Schramski, J. R., Gattie, D. K., Patten, B. C., Borrett, S. R., Fath, B. D., and Whipple, S. J. 2007. Indirect effects and distributed control in ecosystems: distributed control in the environ networks of a seven-compartment model of nitrogen flow in the Neuse River Estuary, USA – Time series analysis. *Ecological Modelling* 206:18–30.

Shevtsov, J., Kazanci, C., and Patten, B. C. 2009. Dynamic environ analysis of compartmental systems: A computational approach. *Ecological Modelling* 220:3219–3224.

Shipley, B. 2002. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press.

Skalski, G. T. and Gilliam, J. F. 2001. Functional responses with predator interference: viable alternatives to the Holling Type II model. *Ecology* 82:3083.

Soil Survey Staff, Natural Resources Conservation Service, U. S. D. o. A. 2012. Web Soil Survey.

Southworth, D., He, X.-H., Swenson, W., Bledsoe, C. S., and Horwath, W. R. 2005. Application of network theory to potential mycorrhizal networks. *Mycorrhiza* 15:589–95.

Spirtes, P., Glymour, C., and Scheines, R. 2001. Causation, Prediction, and Search. MIT Press, Cambridge, MA, 2 edition.

Spirtes, P., Meek, C., and Richardson, T. 1995. Causal inference in the presence of latent variables and selection bias. *In* Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, pp. 499–506. Citeseer.

Stachowicz, J. J. 2001. Mutualism, facilitation, and the structure of ecological communities. *Bioscience* 51.

Stouffer, D. B., Camacho, J., and Amaral, L. a. N. 2006. A robust measure of food web intervality. *Proceedings of the National Academy of Sciences* 103:19015–19020.

Tansley, A. G. 1935. The use and abuse of vegetational concepts and terms. *Ecology* 16:284–307.

Teal, J. M. 1962. Energy flow in the salt marsh ecosystem of Georgia. *Ecology* 43:614–624.

ten Raa, T. 2006. The Economics of Input-Output Analysis. Cambridge University Press, New York.

Terry M Therneau and Beth Atkinson. R port by Brian Ripley. 2010. rpart: Recursive Partitioning.

Toulmin, S. E. 1961. Foresight and understanding: an enquiry into the aims of science. Indiana University Press, Bloomington, IN.

Townsend, C. R., M, R., and Mcintosh, A. R. 1998. Disturbance, resource supply, and food-web architecture in streams. *Ecology Letters* 1:200–209.

Triplehorn, C. and Johnson, N. 2005. Borror and Delong's Introduction to the Study of Insects. Thompson Brooks/Cole, Belmont, CA, 7 edition.

Ulanowicz, R. E. 1997. Ecology, the ascendent perspective. Columbia University Press.

Vandermeer, J. 2006. Oscillating populations and biodiversity maintenance. *BioScience* 56:967–975.

Virgo, N., Law, R., and Emmerson, M. C. 2006. Sequentially assembled food webs and extremum principles in ecosystem ecology. *Journal of Animal Ecology* 75:377–86.

Wardle, D. A., Bardgett, R. D., Klironomos, J. N., Setälä, H., van der Putten, W. H., and Wall, D. H. 2004. Ecological linkages between aboveground and belowground biota. *Science* 304:1629–1633.

Wardle, D. A., Williamson, W. M., Yeates, G. W., and Bonner, K. I. 2005. Trickle-down effects of aboveground trophic cascades on the soil food web. *Oikos* 111:348–358.

Warren, P. H. 1989. Spatial and temporal variation in the structure of a freshwater food web. *Oikos* 55:299–311.

Watts, D. J. and Strogatz, S. 1998. Collective dynamics of 'small-world' networks. *Nature* 393:440–442.

West, G. B., Brown, J. H., and Enquist, B. J. 1997. A general model for the origin of allometric scaling laws in biology. *Science* 276:122–6.

Whipple, S. J. 1998. Path-based network unfolding: A solution for the problem of mixed trophic and non-trophic processes in trophic dynamic analysis. *Journal of Theoretical Biology* 190:263–76.

Whipple, S. J. 1999. Analysis of ecosystem structure and function: extended path and flow analysis of a steady-state oyster reef model. *Ecological Modelling* 114:251–274.

Whipple, S. J. and Patten, B. C. 1993. The problem of nontrophic processes in trophic ecology: Toward a network unfolding solution. *Journal of Theoretical Biology* 163:393–411.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* 26:1–80.

Whittaker, R. H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* 30:279–338.

Williams, R. J., Anandanadesan, A., and Purves, D. W. 2010. The probabilistic niche model reveals the niche structure and role of body size in a complex food web. *PloS One* 5:e12092.

Williams, R. J., Brose, U., and Martinez, N. D. 2006. Allometric scaling enhances stability in complex food webs. *Ecology Letters* 9:1228–36.

Williams, R. J., Brose, U., and Martinez, N. D. 2007. Homage to Yodzis and Innes 1992: scaling up feeding-based population dynamics to complex ecological networks, pp. 37–51. *In* N. Rooney, K. S. McCann, and D. Noakes (eds.), From Energetics to Ecosystems: The Dynamics and Structure of Ecological Systems, chapter 2. Springer, Dordrecht, The Netherlands.

Williams, R. J., Dunne, J. A., and Martinez, N. D. 2002. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences* 99:12917–12922.

Williams, R. J. and Martinez, N. D. 2000. Simple rules yield complex food webs. *Nature* 404:180–183.

Williams, R. J. and Martinez, N. D. 2004. Limits to trophic levels and omnivory in complex food webs: theory and data. *American Naturalist* 163:458–68.

Woodward, G., Ebenman, B., Emmerson, M. C., Montoya, J. M., Olesen, J. M., Warren, P. H., and Valido, A. 2005. Body size in ecological networks. *Trends in Ecology and Evolution* 20:402–9.

Wootton, J. T. 1994. The nature and consequences of indirect effects in ecological communities. *Annual Review of Ecology and Systematics* 25:443–466.

Yodzis, P. 1981. The stability of real ecosystems. *Nature* 289:674–676.

Yodzis, P. 1998. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *Journal of Animal Ecology* 67:635–658.

Yodzis, P. and Innes, S. 1992. Body size and consumer-resource dynamics. *American Naturalist* 139:1151–1175.

Zadeh, L. A. and Desoer, C. A. 1963. Linear system theory: the state space approach. McGraw-Hill, New York.

Ziliak, S. T. and McCloskey, D. N. 2008. The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. University of Michigan Press.