

Michael G. Neubert · Steven C. Blumenshine
Daniel E. Duplisea · Tomas Jonsson · Brenda Rashleigh

Body size and food web structure: testing the equiprobability assumption of the cascade model

Received: 3 February 1999 / Accepted: 26 October 1999

Abstract The cascade model successfully predicts many patterns in reported food webs. A key assumption of this model is the existence of a predetermined trophic hierarchy; prey are always lower in the hierarchy than their predators. At least three studies have suggested that, in animal food webs, this hierarchy can be explained to a large extent by body size relationships. A second assumption of the standard cascade model is that trophic links not prohibited by the hierarchy occur with equal probability. Using nonparametric contingency table analyses, we tested this “equiprobability hypothesis” in 16 published animal food webs for which the adult body masses of the species had been estimated. We found that when the hierarchy was based on body size, the equiprobability hypothesis was rejected in favor of an alternative, “predator-dominance” hypothesis wherein the probability of a trophic link varies with the identity of the predator. Another alternative to equiprobability is that the probability of a trophic link depends upon the ratio of the body sizes of the two species. Using nonparametric regression and likelihood ra-

tio tests, we show that a size-ratio based model represents a significant improvement over the cascade model. These results suggest that models with heterogeneous predation probabilities will fit food web data better than the homogeneous cascade model. They also suggest a new way to bridge the gap between static and dynamic food web models.

Key words Food webs · Body size · Cascade model · Contingency tables · Kernel smoothing

Introduction

Ecologists have reported a number of patterns in food web collections (Lawton 1989; Yodzis 1989; Pimm et al. 1991; Hall and Raffaelli 1993) including patterns in the number of prey per predator (MacDonald 1979); in the proportions of basal, intermediate, and top species (Briand and Cohen 1984; Cohen and Briand 1984); in the length of food chains (Elton 1927; Hutchinson 1959; Pimm and Lawton 1977; Pimm 1982; Briand and Cohen 1987); in the lack of feeding loops (acyclicity) (Gallopini 1972); and in patterns of overlap in prey use by predators (intervality) (Cohen 1977, 1978; MacDonald 1979; Sugihara 1982, 1984). Despite concerns about the quality of the data and the reality and/or relevance of the patterns (e.g., Paine 1988; Hall and Raffaelli 1991; Polis 1991, 1994; Martinez 1993; Polis and Winemiller 1996), the search for a theory to explain the patterns continues (e.g., Bengtsson 1994; Cohen 1994; Martinez and Lawton 1995; Solow 1996; Solow and Beet 1998).

Among the alternative theories (Lawton and Warren 1988; Lawton 1989) are various forms of the phenomenological *cascade model* (Cohen and Newman 1985; Cohen et al. 1990a) which duplicate many of the patterns listed above surprisingly well (Lawton and Warren 1988). The cascade model (which we describe in detail below) makes two crucial assumptions. The first of these is that there is a predetermined trophic hierarchy or “cascade” of the species in any particular food web. This

M.G. Neubert (✉)
Biology Department, MS #34,
Woods Hole Oceanographic Institution,
Woods Hole, MA 02543-1049, USA
e-mail: mneubert@whoi.edu

S.C. Blumenshine
Department of Biological Sciences, Arkansas State University,
State University, AR 72467, USA

D.E. Duplisea
CEFAS Lowestoft Laboratory, Pakefield Rd, Lowestoft,
Suffolk, NR33 0HT, UK

T. Jonsson¹
Department of Biology, Linköping University,
S-58183 Linköping, Sweden

B. Rashleigh
EPA Ecosystems Assessment Branch, 960 College Station Road,
Athens, GA 30605, USA

Present address:

¹ The Rockefeller University, Box 20, 1230 York Avenue,
New York, NY 10021, USA

hierarchy prohibits certain trophic interactions: a given species cannot eat another species that ranks above it in the hierarchy. We say that the cascade model is “phenomenological” because it does not specify the mechanism that generates the hierarchy.

Attempts have been made to put the cascade model on a more mechanistic footing. Perhaps not surprisingly, the approach that has received the most attention focuses on body size, a morphological characteristic that affects almost every aspect of a species’ ecology (Peters 1983; Calder 1984). Body size is correlated with metabolic rates, ingestion rates, vital rates, locomotive ability, home range size, and (negatively) with population density to name but a few examples. Elton (1927) noted that body size also constrains the size of food that a predator can eat: consumers are generally larger than what they consume. This pattern has been documented in both aquatic (Sheldon et al. 1972, 1973; Hansen et al. 1994) and terrestrial (Vézina 1985) habitats, for both vertebrate (Gittleman 1985; Vézina 1985) and invertebrate (Warren and Lawton 1987) predators, and across a range of metabolic types (Cohen et al. 1993). There are exceptions to the rule—parasites, decomposers, and pack hunters for example—but in documented food webs these are indeed exceptions. In an examination of 31 published food webs, Cohen et al. (1993) determined that the predator was larger than the prey in nearly 90% of the trophic links.

Warren and Lawton (1987), Cohen (1989a), and Cohen et al. (1993) have suggested that body size provides a mechanistic interpretation of the assumed trophic hierarchy of the cascade model, albeit an imperfect one. Cohen et al. (1993) showed that when species are ranked by body size, webs conform more closely to a trophic hierarchy than would be expected if ranks were assigned randomly.

The second crucial assumption of the cascade model is that within a web, the trophic links that are not prohibited by the hierarchy are assumed to be independent and identical Bernoulli random variables. In other words, the probability that a species eats some species below it in the hierarchy is assumed to be the same for any such pair of species. We call this the *equiprobability assumption*.

But do body-size-based rankings agree with the equiprobability assumption of the cascade model? After reviewing the cascade model and its variants in the next section, we test the equiprobability assumption using food webs in which the predators are larger than their prey and in which the species have been ranked by body mass. The tests that we use are equivalent to tests for homogeneity of proportions in contingency tables. Our analysis shows that predation probabilities tend to differ among the predators in a web, but seem to be constant among the prey. Thus the probability of a trophic link between any two species should be more related to, for example, the hunting abilities or morphology of the predator than, for example, the abundance or defensive abilities of the prey.

Although there is less evidence that prey sizes are also constrained from below, it is reasonable to suspect

that prey that are too small might not contain enough energy to justify their capture and consumption. As a result, there might be a preferred prey size range that is a function of the predator’s size (Thiebaut and Dickie 1993; Hansen et al. 1994). Because predator-prey pairs are usually part of a complex food web, it is also possible that predator-prey size relationships are affected by dynamic constraints working on the whole community, perhaps via selective extinctions (Fowler and MacMahon 1982) of predators with extremely large or small prey. Jonsson (1998) showed that predator-prey size ratios decrease, on average, with trophic height. Jonsson and Ebenman (1998) showed that a correlation between interaction strengths and predator-prey size ratios in a food chain model tended to decrease the characteristic return time to equilibrium. Do these constraints generate a non-uniform frequency distribution of predator-prey size ratios within a community? To find out, we employed a nonparametric regression procedure called *kernel smoothing*. We describe this method and the results of our analysis in the penultimate section. It indeed turns out that the predation probability as a function of predator-prey size ratio can be significantly different from constant.

We describe the food web data we have used in the next section. Like any data, they have shortcomings. Some of these have been addressed by other authors (see reference above); some we address in the concluding discussion section. Despite these deficiencies, our analyses reveal evidence that the webs that we studied are structured in a biologically reasonable way. Our results also suggest important ways to improve the biological realism of food web models.

The data

The webs that we used for this study are a subset of 53 webs from the EcoWeb database (Cohen 1989b). Estimates of average adult body mass have been made for most of the species in these webs. The webs were edited as follows: (1) Whenever present, humans were omitted. (2) Since it is difficult (perhaps impossible) to assign body masses to them, producers (or basal species) were also omitted. The webs thus became “consumer webs.” (3) Body mass estimates were not available for some species; these species were deleted. (4) Body mass estimates for trophic species were obtained by averaging the estimated adult body masses of the constituent biological species. (5) Thirty-seven webs had at least one predator which consumed a prey larger than itself and thus had to be eliminated from our analysis. The remaining 16 webs—which we use in the following analyses—are listed in Table 1 along with their sizes before and after manipulation and the number of links in the resulting consumer webs.

The body mass data for webs 2–39 were provided by P. Yodzis (unpublished data). They were previously used by Yodzis (1984) to study the effect of metabolic type on

Table 1 Potentially size-structured food webs. Web number represents the position of the web in the EcoWeb database (Cohen 1989b; Cohen et al. 1990a). *Consumers* are the number of species after elimination of basal species, species without body mass estimates, and humans. Values in parentheses indicate the number of species omitted due to lack of body mass data

Web number	Habitat description	Species	Consumers	Links
2	Knysna estuary, South Africa	15	12	15
5	Salt marsh, Georgia	7	4	2
14	Mangrove swamp 1, Hawaii	8	7	6
16	Pamlico estuary, North Carolina	14	10	11
20	Antarctic pack ice zone	19	16	24
23	Prairie, Manitoba	15	14	19
36	Mangrove swamp 6, Hawaii	22	14	10
39	Lake Nyasa, sandy shore, Malawi	37	32	43
42	Upwelling areas, Pacific Ocean	16	11	31
49	Nearshore marine 2, Aleutian Islands	12	8 (1)	7
54	Western rocky shore, Barbados	15	8 (3)	6
61	Barren regions, Spitsbergen	9	7	6
62	Reindeer pasture, Spitsbergen	12	9	7
65	Yoshino River rapids, Japan	16	12 (2)	22
66	River Thames, England	10	7	10
68	Loch Leven, Scotland	22	15 (3)	15

predator-prey interactions and by Cohen et al. (1993) to examine the effect of body size on food web structure. One of us (T. Jonsson) estimated the body masses of the species in webs 42–68.

Stochastic food web models: the cascade model and heterogeneous alternatives

A convenient way to summarize the feeding relationships among n species in a hypothetical food web is an $n \times n$ predation matrix: if species j eats species i , then make the ij th element of the matrix a “1”; otherwise make it a “0”. One can imagine that for an actual food web, the predation matrix is a realization of an $n \times n$ matrix random variable \mathbf{A} .

The cascade model is a stochastic process that generates the distribution of \mathbf{A} . Each element of \mathbf{A} is assumed to be an independent binary random variable with $\Pr(A_{ij} = 1) = p_{ij}$. The predation probabilities p_{ij} are determined by an (unknown) ranking of the species within the community. If the rank of species j is less than or equal to the rank of species i , $p_{ij} = 0$. If the rank of species j is greater than the rank of species i , then $p_{ij} = p$, p a constant. By naming the species according to their rank (which we do, without loss of generality, from here on) the cascade model predation probabilities satisfy

$$\text{cascade model: } p_{ij} = \begin{cases} p & \text{if } i < j, \\ 0 & \text{if } i \geq j. \end{cases} \quad (1)$$

Any realization of \mathbf{A} under the cascade model will be strictly upper triangular (i.e., every element of the realization laying on or below the main diagonal will be zero). This structure excludes the possibility of cannibalism (since $p_{ii} = 0$) as well as trophic cycles (e.g., i eats j , j eats k , and k eats i). Finally, we note that there may be more than one ranking which produces a strictly upper triangular predation matrix representation for a given food web. All such rankings are called *admissible*.

Cohen (1990) introduced thirteen alternatives to the cascade model that relax the equiprobability assumption

$p_{ij} = p$, for $i < j$. Each of these alternatives falls into one of three categories: prey dominant, predator–dominant, or distance dominant. Other assumptions—concerning the expected number of prey for a given predator (or vice versa), and on the relationship between the number of species and the expected number of trophic links—distinguish models within the three categories.

In a prey-dominant model, the predation probabilities are the same for every potential predator of a particular prey species (i.e., for every species of higher rank than the prey species). On the other hand, qualities intrinsic to the various prey species—such as abundances or defensive abilities—may produce different predation probabilities among the potential prey of a particular predator. Thus the predation probabilities only depend upon the identity of the prey species:

$$\text{predator - dominant model: } p_{ij} = \begin{cases} p_j & \text{if } i < j, \\ 0 & \text{if } i \geq j. \end{cases} \quad (2)$$

Prey-dominant models assume that the predation probabilities are constant within rows of (the upper triangular part of) \mathbf{A} , but may differ between rows.

The situation is reversed in predator-dominant models, in which the predation probabilities only depend upon the identity of the predator. Every potential prey of a given predator is equally likely to be part of that predator’s diet, but the predation probabilities may differ between predators because of differences, for example, in hunting abilities or morphology. Predator dominance implies that the predation probabilities are constant within columns of \mathbf{A} , but may differ between columns:

$$\text{predator - dominant model: } p_{ij} = \begin{cases} p_j & \text{if } i < j, \\ 0 & \text{if } i \geq j. \end{cases} \quad (3)$$

If a species’ rank correlates with its body size, then differences in body size between consumers and their prey might determine the predation probabilities. A third class of alternatives, distance-dominant models, tries to capture this structure by assuming that the predation probabilities depend upon the difference between the ranks of the predator and the prey:

$$\text{distance - dominant model: } p_{ij} = \begin{cases} p_{j-i} & \text{if } i < j, \\ 0 & \text{if } i \geq j. \end{cases} \quad (4)$$

Distance-dominant models have predation probabilities that are constant along diagonals of **A**, but may differ between diagonals. (Note, however, that a small difference in rank does not necessarily translate into a small difference in body size.)

Testing the equiprobability assumption: predation matrices as contingency tables

In this section we describe a method for testing whether the food webs described above are consistent with the equiprobability assumption of the cascade model. The method is based on a standard test for the homogeneity of proportions in contingency tables (Everitt 1977; Collett 1991).

We begin with an example. Consider a modified version of web 42 (Fig. 1), and its 11×11 predation matrix

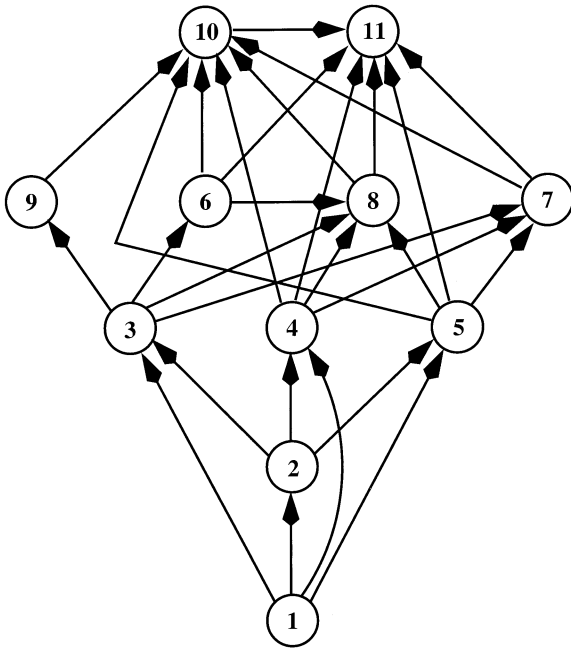


Fig. 1 Consumer web 42, upwelling areas, Pacific Ocean. The species, ranked by increasing size, are 1 bacteria; 2 zooflagellates; 3 ciliates; 4 small calanoids; 5 meroplankton, appendicularians, and doliolids; 6 medium-sized calanoids; 7 cyclopoids; 8 calanoids, small tomopterids, and small coelenterates; 9 juvenile euphausiids; 10 chaetognaths and polychaetes; 11 anchovy

Table 2 A contingency table for web 42, upwelling areas, Pacific Ocean (Fig. 1). The table was constructed by assuming that each column is a separate sample

	Column number (predator rank)									
	2	3	4	5	6	7	8	9	10	11
Successes	1	2	3	3	1	3	5	1	6	6
Failures	0	0	0	1	4	3	2	7	3	4

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ & & & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ & & & & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ & & & & & 1 & 0 & 1 & 1 & 0 & 0 \\ & & & & & & 0 & 1 & 1 & 0 & 0 \\ & & & & & & & 1 & 0 & 1 & 0 \\ & & & & & & & & 1 & 0 & 0 \\ & & & & & & & & & 1 & 0 \\ & & & & & & & & & & 1 \end{pmatrix}. \quad (5)$$

(The blank spaces in the matrix are all “0”s.) Each row (or column or diagonal) of the upper triangular part of the matrix can be thought of as a separate sample. In matrix (5), counting by columns, the first sample has only one observation, the second sample has two, and so on until the tenth and last sample which has ten observations. As is standard in the statistical literature on binary data, we call the 1s “successes” and the 0s “failures.” These data can then be summarized in the form of a contingency table (Table 2), which presents the number of successes and failures in each of the *m* samples. The number of samples is always one less than the number of species: *m* = *n*−1.

The cascade model assumes that the predation probability for the *i*th sample, *p_i*, is the same for every sample and equal to *p*. We call this null hypothesis the *equiprobability hypothesis*:

$$H_e: p_1 = p_2 = \dots = p_m = p. \quad (6)$$

An estimate of *p*, call it \hat{p} , can be calculated as the total number of successes divided by the total number of successes plus failures. (\hat{p} is also called the *connectance* of a web.)

The alternative to the hypothesis is that at least one of the samples has a different predation probability than the others. The interpretation of this alternative hypothesis depends upon how we define the samples. If we define samples as rows in the predation matrix, then the alternative hypothesis is equivalent to:

$$H_{\text{prey-dominance}}: \text{predation probabilities differ between rows,} \quad (7)$$

the assumption of a prey-dominant model. Similarly, we could take columns in the predation matrix as samples and get the alternative hypothesis

$$H_{\text{predator-dominance}}: \text{predation probabilities differ between columns,} \quad (8)$$

or take diagonals and obtain

$$H_{\text{distance-dominance}}: \text{predation probabilities differ between diagonals.} \quad (9)$$

Table 3 Testing for homogeneity of proportions. The significance level is based on 10,000 randomizations

Web number	Habitat description	Significance level (P)			
		\hat{p}	Prey dominant	Predator dominant	Distance dominant
2	Knysna estuary, South Africa	0.23	0.94	0.12	0.37
5	Salt marsh, Georgia	0.33	0.60	0.40	0.61
14	Mangrove swamp 1, Hawaii	0.29	0.93	0.59	0.35
16	Pamlico estuary, North Carolina	0.24	0.56	0.42	0.46
20	Antarctic pack ice zone	0.20	0.37	0.61	0.81
23	Prairie, Manitoba	0.21	0.18	0.18	0.76
36	Mangrove swamp 6, Hawaii	0.11	0.46	0.06	0.34
39	Lake Nyasa, sandy shore, Malawi	0.09	<0.01	<0.01	0.35
42	Upwelling areas, Pacific Ocean	0.56	0.80	0.06	0.11
49	Nearshore marine 2, Aleutian Islands	0.25	0.93	0.67	0.03
54	Western rocky shore, Barbados	0.21	0.38	0.30	0.76
61	Barren regions, Spitsbergen	0.29	0.14	0.24	0.80
62	Reindeer pasture, Spitsbergen	0.19	1.00	<0.01	0.95
65	Yoshino River rapids, Japan	0.33	0.33	<0.01	0.30
66	River Thames, England	0.48	0.59	0.09	0.80
68	Loch Leven, Scotland	0.14	0.28	0.05	0.48

A standard method for testing the null hypothesis H_e against one or another of the above alternatives uses the X^2 -statistic (Pearson 1904). This statistic measures the extent to which observed counts in the contingency table deviate from those expected under the equiprobability hypothesis. Let E_i and O_i , respectively, represent the expected and observed number of successes out of the m_i observations in the i th column of the contingency table. Similarly let e_i and o_i be the expected and observed number of failures, respectively. Under the equiprobability hypothesis, $E_i = \hat{p} m_i$ and $e_i = (1 - \hat{p}) m_i$. Finally, X^2 is defined by the formula:

$$X^2 = \sum_{i=1}^m \frac{(O_i - E_i)^2}{E_i} + \frac{(o_i - e_i)^2}{e_i}. \quad (10)$$

Large values of X^2 (relative to its distribution) signify large deviations from expected counts and suggest rejection of the cascade hypothesis in favor of one of its alternatives. With a large number of observations in each sample, the distribution of X^2 under the equiprobability hypothesis is approximately χ^2 on $m-1$ degrees of freedom (Everitt 1977). However, the number of observations in the first few columns of a web (or the last few rows or diagonals) is always small. As a result, the χ^2 approximation may not be very accurate. (The problem is exacerbated in webs with few species.) In addition, the support of the distribution includes both matrices with no links and matrices with every possible link! We would prefer to have the distribution of X^2 over all the predation matrices of the same size and with the same number of links.

To avoid these potential pitfalls, we generated an approximation to the conditional distribution of X^2 given the same number of links by iteratively randomizing the locations of 1's in the upper triangular portion of the observed predation matrix and calculating X^2 for each randomization. The attained significance level (P) is the fraction of the randomizations with X^2 values greater

than or equal to the observed value. Continuing with our example, $X^2 \approx 15.34$ and $P \approx 0.06$ for Table 1, suggesting that the predation probabilities differ between the columns of web 42.

We applied this contingency table method to the 16 food webs described earlier. Table 3 displays our results. While there is essentially no evidence for prey or distance dominance in these webs, there is evidence for predator dominance: H_e was rejected in favor of $H_{\text{predator-dominance}}$ at a significance level of $P \leq 0.06$ in 6 of the 16 webs (numbers 36, 39, 42, 62, 65, and 68). The probability of incorrectly rejecting the equiprobability hypothesis six times at this significance level is approximately 0.00002. The evidence is overwhelming for three webs (numbers 39, 62, and 65), with $P \leq 0.005$.

Testing for nonuniform body size ratios: smoothing the binary data of a food web

Each of the stochastic food web models considered above assumes that the probability of a trophic link between two species depends only upon their relative ranks on some scale. A reasonable alternative, and the object of our scrutiny in this section, is that the predation probabilities depend more specifically on their actual *position* on the scale.

Suppose we assign the rank of species i based upon its size, s_i , such that $s_i < s_j$ if $i < j$. We then posit two alternative models for the predation probabilities p_{ij} : the cascade model (Eq. 1) and the size-ratio model,

$$\text{size-ratio model: } p_{ij} = \begin{cases} f(\log[s_j / s_i]) & \text{if } i < j, \\ 0 & \text{if } i \geq j. \end{cases} \quad (11)$$

In this section, we describe and implement a method for estimating the predation probability function f , for each of the size-ranked food webs of Table 2. Because the size-ratio model contains the cascade model as a special

case, it fits the data at least as well as the cascade model. We therefore use likelihood ratio tests to determine whether the improvement, if any, is statistically significant.

We take the domain of the predation probability function f to be the closed interval between the smallest and largest log size ratios. (Since f is a probability, its range is $[0,1]$.) The data available for estimating f are binary: for the ij th entry in the predation matrix (with $i < j$) we have the corresponding size ratio s_j/s_i and either an observation of a trophic link (a “1”) or the lack thereof (a “0”).

A simple way to construct a step-function estimate of f is to break up its domain into nonoverlapping line segments and approximate f on each segment by the fraction of the observations in that segment that are links (i.e., build a regressogram; Tukey 1961). The resulting estimate can be very jagged and its shape is sensitive to the subjectively defined length and position of the segments. We therefore opt for a different method, called *kernel smoothing*, that produces a smooth curve. Comprehensive treatments of this nonparametric regression method can be found in Silverman (1986), Härdle (1990), and Hastie and Tibshirani (1990), but we will use a few paragraphs here to review the basics.

Let $x_i, i=1, \dots, n(n-1)/2$, represent the logarithms of the size ratios of the permissible predator-prey pairs. For each x_i , let y_i equal 1 if there is a trophic link between the predator and prey which produced the ratio x_i and equal 0 otherwise. We can produce an estimate of f at any size ratio x by taking a weighted average of the data, assigning the weight to a particular data point based on its distance from x . The function which produces the weight is called the kernel, K . The resulting estimate, $\hat{f}(x;h)$, is given by the formula (Nadaraya 1964; Watson 1964)

$$\hat{f}(x;h) = \frac{\sum y_i K((x-x_i)/h)}{\sum K((x-x_i)/h)}. \tag{12}$$

K is typically a positive symmetric nonincreasing function; thus the weights decrease with distance from x . The parameter h , called the *bandwidth*, controls the rate of decrease in weight with distance. Because this procedure produces a continuous curve, \hat{f} is called a “smooth” of the data.

There are two choices we must make to implement this kernel-smoothing procedure: which kernel and what bandwidth. We will use the Epanechnikov kernel

$$K_e(u) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5}\right), & -\sqrt{5} \leq u \leq \sqrt{5}, \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

(Epanechnikov 1969) because it has certain optimality qualities (Gasser et al. 1985; Härdle 1990). Note that because K_e is positive over only a finite range, some choices of x in Eq. 12 can result in the undefined quantity “0/0.” When this occurs we define $\hat{f}(x;h) = 0$ at that value of x .

In practice, the choice of the kernel is less important than the choice of the bandwidth, which determines the

amount of smoothing. A value of h that is too small results in very little smoothing and can produce spurious small-scale structure in the smooth. On the other hand, a value that is too large oversmooths the data, obscuring important features of the true curve f . In fact by Eq. 12, in the limit as $h \rightarrow \infty, \hat{f} \rightarrow \hat{p}$, the estimate of the predation probability function of the cascade model. Thus the cascade model is a special case of the size-ratio model.

Figure 2 shows how the bandwidth (reported as a percentage of the range of the data) affects the weights assigned to the data. The dots represent every potential predator-prey pair in web 42. Their abscissas are the logarithms of the ratios of predator size to prey size. Their ordinates are the weights they are assigned by Eqs. 12 and 13 when estimating f at a point in the middle of its range ($x \approx 7$). Combining Eqs. 12 and 13, it is evident that data falling outside a neighborhood of width $2h\sqrt{5}$ centered on x receive zero weight. As h increases, more data are given positive weights. (Note that the weights always sum to 1, although it may not appear so in the left panel of Fig. 2. There are four predator-prey pairs with size ratios whose logarithm is very nearly 7. These four points are so close together that they appear as a single dot at the peak.) Figure 3 displays the smooths of web 42 that result from these bandwidths.

For data summary, a subjective choice may be preferable to other methods for selecting the bandwidth. Here, however, we are generating the smooth for inferential purposes and therefore adopt an automatic method: likelihood cross-validation (Habbema et al. 1974; Stone 1974; Duin 1976).

The idea is simple. Assume that we obtained, in addition to the data that we already have, one more observation Y at size ratio X . The log likelihood of Y given our estimate \hat{f} derived from the earlier data would then be given by:

$$L(h) = \log[Y\hat{f}(X;h) + (1-Y)(1-\hat{f}(X;h))], \tag{14}$$

and we could choose h to maximize L . Because we do not have this extra datum, we instead leave out the i th observation from the data and treat it as if it was our new observation. We compute a new estimate of the predation probability function based on the reduced data set (\hat{f}_i) and then calculate the log likelihood of the left out observation:

$$L_i(h) = \log[y_i\hat{f}_i(x_i;h) + (1-y_i)(1-\hat{f}_i(x_i;h))]. \tag{15}$$

There was nothing special about the i th observation—we could have left any of them out—so we average over all of the data to obtain:

$$CV(h) = \frac{2}{n(n-1)} \sum_{i=1}^{n(n-1)/2} L_i(h). \tag{16}$$

The likelihood cross-validation choice of h, h^* , is the one which maximizes CV .

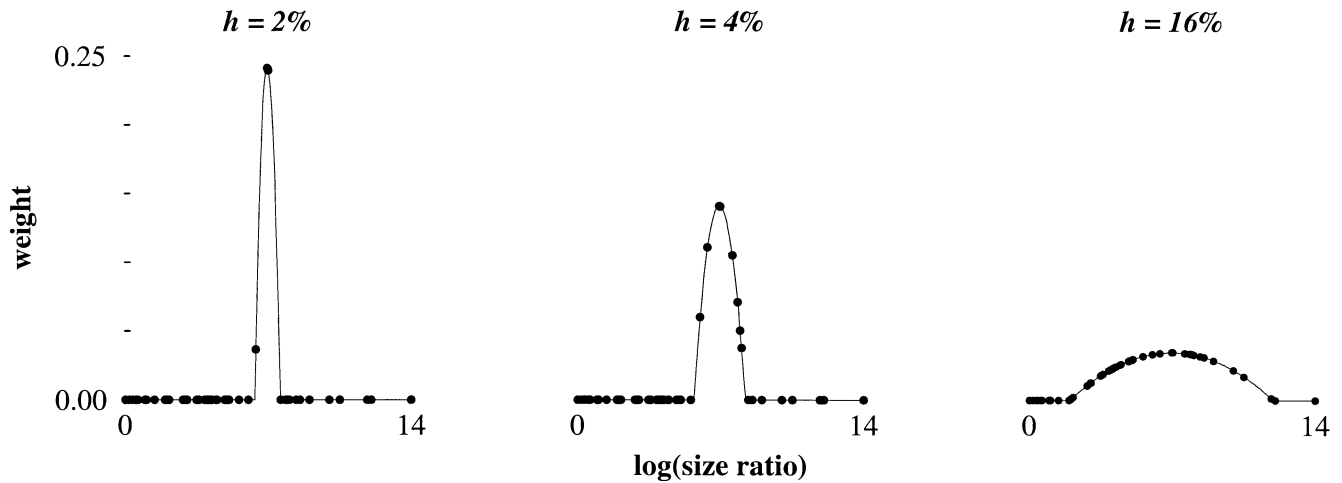


Fig. 2 The effect of bandwidth choice on the weighting of data. The dots represent all the possible predator-prey pairs in web 42 (cf. Fig. 1). Their ordinates give the weight that each point is given when forming an estimate of the predation probability function $\hat{f}(x;h)$ at the $\log(\text{size ratio})$ $x \approx 7$. The bandwidth h is shown as a percent of the maximum $\log(\text{size ratio})$. As h increases, more data are weighted and the weight per data point decreases

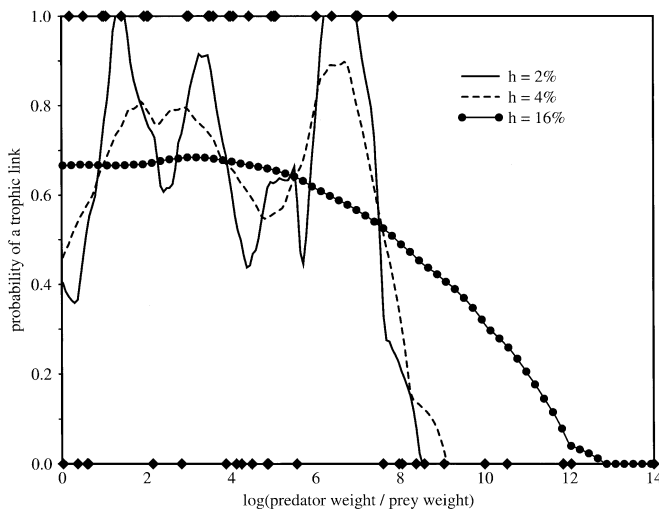


Fig. 3 The effect of bandwidth (h) on the estimated predation probability function $\hat{f}(x;h)$ of the size-ratio model for web 42. Small values of the bandwidth cause \hat{f} to fluctuate wildly (*solid curve*), while large values (*dotted line*) tend to “oversmooth” the curve. An automatic procedure for choosing the bandwidth based on likelihood cross-validation (described in the text) in this case produces an intermediate value of h (*dashed curve*)

The data for each of the webs of Table 2 are displayed in Fig. 4 (diamonds), along with the estimated probability functions under the size-ratio model obtained by smoothing (solid curves), and the estimated probability functions under the cascade model (dashed lines). The shaded areas represent ± 2 SEs at each point (see Copas 1983 for their computation). The bandwidths used to produce the smooths are listed in Table 4.

In some cases, \hat{f} closely approximates the constant \hat{p} . For other webs (eg., 5, 39, 42, 61, 62, and 65), the smooths look nothing like straight lines. But do these smooths generated by the size-ratio model represent significantly better fits to the data than the constant estimates of the cascade model? To find out, we generated a *likelihood ratio statistic* (see e.g., Mendenhall et al. 1986), Λ , which compares the likelihood of the data under the two models:

$$\Lambda = 2 \log \left\{ \frac{\prod [y_i \hat{f}(x_i; h^*) + (1 - y_i)(1 - \hat{f}(x_i; h^*))]}{\prod [y_i \hat{p} + (1 - y_i)(1 - \hat{p})]} \right\}. \quad (17)$$

Large values of Λ , relative to its distribution, suggest that the size-ratio model fits the data significantly better than the cascade model.

We approximated the distribution of Λ by iteratively randomizing the sequence $\{y_i\}$, generating the likelihood cross-validated estimate \hat{f} for each randomized sequence, and then computing Λ for each \hat{f} . The fraction of the randomizations with values of Λ greater than or equal to the observed value represent an approximate attained significance level (P -value). As in the contingency table analysis, this randomization procedure has the advantage of limiting support of the distribution of Λ to predation matrices with the same number of links occurring at the same size ratios as the observed predation matrix.

The size-ratio model was a significant improvement over the cascade model ($P < 0.05$) in four cases (Table 4). (The probability of falsely rejecting the cascade model four times at this level is roughly 0.006.) The contingency table analysis distinguished these same webs (39, 42, 62, and 65) as being predator dominated. For webs 14, 23, 54, and 66, $h^* \rightarrow \infty$ and $\hat{f}(x; h^*) \rightarrow \hat{p}$. Although $\Lambda = 0$ by necessity for these webs, this value is not significantly small: $h^* \rightarrow \infty$ for at least half of the randomizations of these webs as well (i.e., a large value of h^* does *not* provide evidence that the cascade model fits better than the size-ratio model).

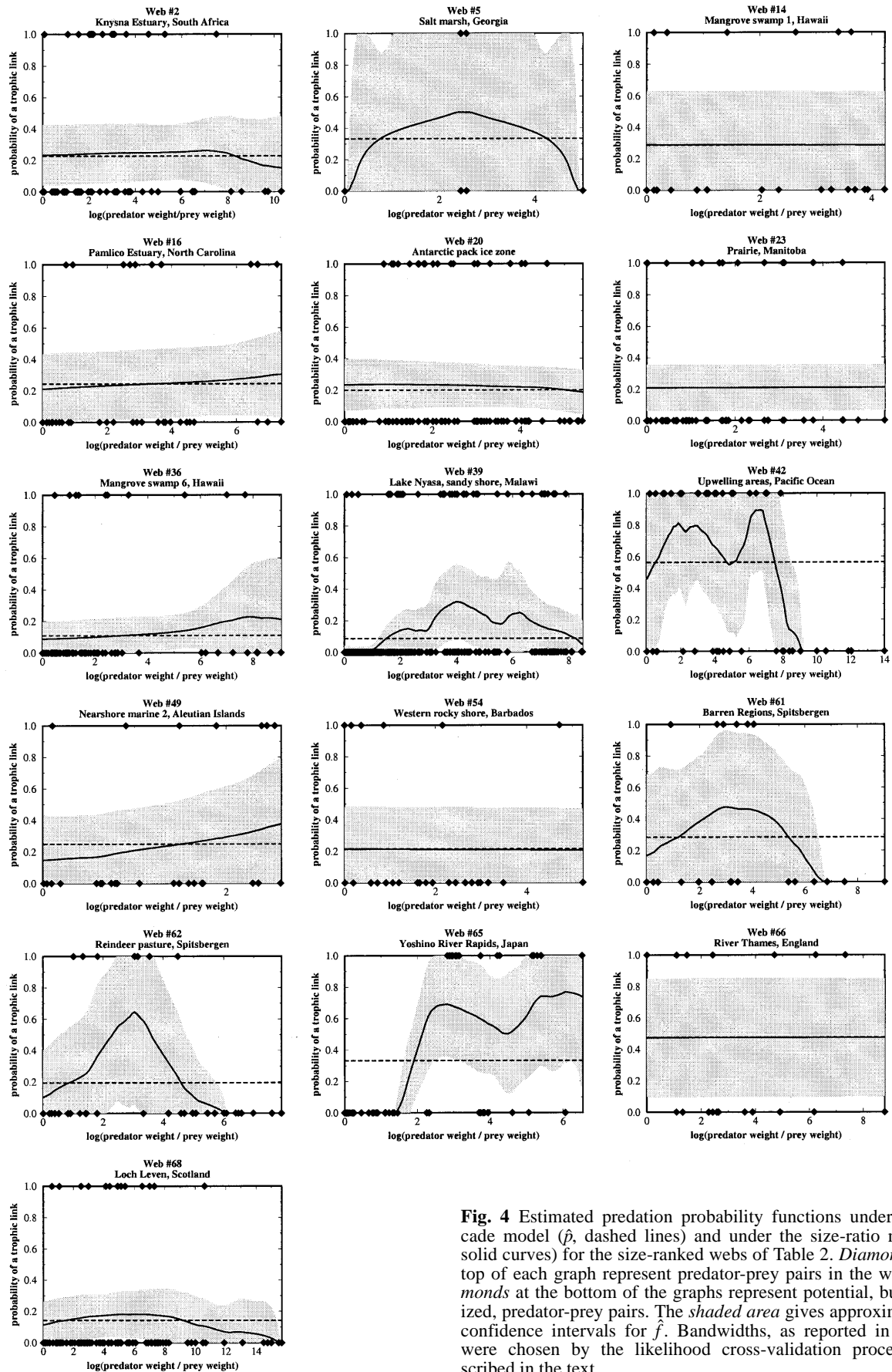


Fig. 4 Estimated predation probability functions under the cascade model (\hat{p} , dashed lines) and under the size-ratio model (\hat{f} , solid curves) for the size-ranked webs of Table 2. *Diamonds* at the top of each graph represent predator-prey pairs in the webs. *Diamonds* at the bottom of the graphs represent potential, but unrealized, predator-prey pairs. The *shaded area* gives approximate 95% confidence intervals for \hat{f} . Bandwidths, as reported in Table 4, were chosen by the likelihood cross-validation procedure described in the text

Table 4 Testing for nonuniform predator-prey size ratios. h is reported as a fraction of the maximum log size ratio. P is based on 1000 randomizations

Web number	Habitat description	h	Λ	P
2	Knysna estuary, South Africa	0.28	0.41	0.44
5	Salt marsh, Georgia	0.21	0.91	0.26
14	Mangrove swamp 1, Hawaii	∞	0.00	1.00
16	Pamlico estuary, North Carolina	0.42	0.39	0.39
20	Antarctic pack ice zone	0.21	1.05	0.29
23	Prairie, Manitoba	∞	0.00	1.00
36	Mangrove swamp 6, Hawaii	0.29	0.76	0.34
39	Lake Nyasa, sandy shore, Malawi	0.05	24.28	<0.01
42	Upwelling areas, Pacific Ocean	0.04	9.92	<0.01
49	Nearshore marine 2, Aleutian Islands	0.31	0.98	0.27
54	Western rocky shore, Barbados	∞	0.00	1.00
61	Barren regions, Spitsbergen	0.13	2.39	0.16
62	Reindeer pasture, Spitsbergen	0.09	5.25	0.04
65	Yoshino River rapids, Japan	0.10	19.05	<0.01
66	River Thames, England	∞	0.00	1.00
68	Loch Leven, Scotland	0.14	1.23	0.29

Discussion

Using two statistical approaches—contingency tables and kernel smoothing—we have shown that the predation probabilities of the cascade model can be significantly different from constant when the species are ranked by body mass.

Our contingency table analysis suggests that a predator-dominant model fits the data better than the cascade model. This finding complements that of Cohen (1990), who evaluated 13 heterogeneous alternatives to the cascade model. Cohen considered the “predator-dominant, constant-column-sum, linear-link-scaling model” to be the only viable alternative to the cascade model. In this model, the predation probabilities decrease with increasing predator rank in a way that holds the expected number of prey per predator constant. The model predicts the proportions of basal, intermediate, and top species in the 113 EcoWeb community food webs better than the cascade model. It also more closely predicts the proportions of basal-top and intermediate-top links. The cascade model, however, is better at predicting the proportion of intermediate-intermediate links, the proportions of basal-intermediate links, and the observed modal chain length. While the contingency table method that we used does not indicate *how* the predation probabilities differ from constant, such methods exist. It would be interesting to see whether the predation probabilities do in fact decrease with predator rank as the predator-dominant, constant-column-sum model requires. (A preliminary investigation does not suggest any clear pattern.)

Our kernel smoothing analysis showed that the probability of an actual trophic link between a potential predator-prey pair depends upon the ratio of the size of the (larger) predator to the size of the (smaller) prey. Again, our analysis does not indicate how these predation probabilities differ from constant, only that they do. However, three graphs in Fig. 4 for which the null hypothesis of constant probability was rejected (39, 42, 62), along with three of those for which it was not (5, 61, 68), do suggest a pattern: a peak in predation proba-

bility at an intermediate predator-prey size ratio. As we pointed out in the introduction, this is the pattern one might expect based on energetic and anatomical constraints.

These analyses represent a second step in the process of building from the phenomenological cascade model to a more mechanistic model. The first step was to demonstrate that body size can be used as an approximation to the hierarchy of the cascade model (Warren and Lawton 1987; Cohen et al. 1993). Here we have shown that when body size *is* used to generate the hierarchy, the cascade model assumption of constant predation probability should be replaced either with predator-dominant predation probabilities or with predation probabilities that depend upon the size ratio between the predator and its prey.

Our results also help to bridge the gap between models that describe the static properties of food webs and those that describe food web dynamics. The static properties of a food web—the connectance, the length of food chains, the degree of omnivory etc.—are derived from the feeding relationships between its component species. These are the properties that the cascade model and its variants attempt to predict. The cascade model does not predict anything about the dynamics of a food web—the way the population sizes of the species change through time. Lotka-Volterra models, composed of systems of coupled differential equations, are typically used for this purpose (May 1973). Lotka-Volterra models, however, say nothing about how trophic interactions within a web are determined.

In real communities, both static and dynamic properties of food webs play a role in determining structure, function, and stability. It is therefore useful—from both practical and theoretical perspectives (Caswell 1988)—to have models that capture both aspects. Recognizing this fact, Cohen et al. (1990b), coming from the static end of the spectrum, developed a hybrid Lotka-Volterra cascade model, and used it to address the relationship between the number of species in a food web and its stability. In essence, the model assigns trophic links using the homo-

geneous version of the cascade model. It then uses another stochastic process to determine whether each link has a positive or zero effect on the population growth rate of the predator and a negative or zero effect on the population growth rate of the prey. The Lotka-Volterra cascade model does not, however, assign the magnitude of these effects. Coming from the other end of the spectrum, Yodzis and Innes (1992) developed a technique to assign numerical values to the parameters in dynamic food web models based on body size, energetics, and allometry. They used their model to study the stability of a two-species predator-prey model. Jonsson and Ebenman (1998) use a similar approach to study the stability of food chains. The Yodzis-Innes method, however, is silent on the issue of how trophic links are established.

Our results suggest a logical next step: combining a log-normal distribution of body sizes (Cohen et al. 1993); a predator-dominant, body-size-based version of the Lotka-Volterra cascade model to generate trophic structure; and a Yodzis-Innes-type allometric model to determine the magnitudes of the parameters. Analysis of such a model would advance our understanding of community dynamics by constraining both the static and dynamic properties of the model webs in ways that are consistent with empirical data.

Several caveats have no doubt crossed the reader's mind, and we would be remiss if we did not recognize them here. First, we remind the reader that 70% of the webs that we tested had to be eliminated from our analyses because they had at least one predator that was smaller than its prey. Body size alone cannot determine the trophic hierarchy in these communities. Nevertheless, body size may be a good approximate ranking for these webs (Cohen et al. 1993). Indeed it would be interesting to see if the body size rankings of these webs are close to the set of admissible rankings. The distance between two rankings might be measured by the minimum number of two-species, one-rank swaps required to bring one ranking into agreement with the other.

Our other reservations stem from the variable resolution of documented food webs. First, some webs list "trophic species" formed by lumping together several biological species. The average of the body sizes of the constituent species may not be close to the body size of any of the individual species. Second, even when biological species are reported, they are not typically resolved into size classes. Since only adult body sizes were estimated, individual growth and ontogenetic niche shifts cannot be accounted for. Artfactual predator-prey size relationships may result. For example, consider a northern lake food web, where fish species are known to change their preferences for both prey species and prey body size as they develop (Bergman and Greenberg 1994). Small individuals of a predatory fish species may feed heavily on zooplankton, but become strictly piscivorous as they grow. Both trophic links might be recorded in the food web, but only the adult body size will be assigned to the predator. This produces an error because there is never a trophic link between the adult-

sized fish and zooplankton. Similar errors can occur if a predator consumes only the juvenile form of some prey species, but the adult prey size is recorded.

Despite these deficiencies in the data, our analyses reveal evidence that the webs that we studied are structured in biologically reasonable ways. When better food web data sets become available, the methods we have presented here will provide useful tools for their analysis.

Acknowledgements This paper was improved by discussions with the participants in the summer workshop, "Food Web Structure and Dynamics in Marine, Terrestrial and Freshwater Environments," held at Cornell University and sponsored by the Office of Naval Research and Woods Hole Oceanographic Institution. In particular, it is a pleasure to acknowledge Andy Solow, Hal Caswell and Joel Cohen for their comments, criticisms, and advice. We thank Peter Yodzis for providing body mass estimates for a number of webs and Bob Holt and C.H. Peterson for reviewing the manuscript. Finally, we thank the Cornell Theory Center for the use of their computing facilities during the workshop. WHOI Contribution No. 10056.

References

- Bengtsson J (1994) Confounding variables and independent observations in comparative analyses of food webs. *Ecology* 75: 1282–1288
- Bergman E, Greenberg LA (1994) Competition between a planktivore, a benthivore and a species with ontogenetic diet shifts. *Ecology* 75:1233–1245
- Briand F, Cohen JE (1984) Community food webs have scale-invariant structure. *Nature* 307:264–267
- Briand F, Cohen JE (1987) Environmental correlates of food chain length. *Science* 238:956–960
- Calder WA (1984) Size, function, and life history. Harvard University Press, Cambridge, MA
- Caswell H (1988) Theory and models in ecology: a different perspective. *Ecol Model* 43:33–44
- Cohen JE (1977) Food webs and the dimensionality of trophic niche space. *Proc Natl Acad Sci USA* 74:4533–4536
- Cohen JE (1978) Food webs and niche space. Princeton University Press, Princeton, NJ
- Cohen JE (1989a) Food webs and community structure. In: Roughgarden J, May RM, Levin SA (eds) *Perspectives in ecological theory*. Princeton University Press, Princeton, NJ, pp 181–202
- Cohen JE (1989b) Ecologists' co-operative web bank, version 1.0. Machine-readable database of food webs. Rockefeller University, New York
- Cohen JE (1990) A stochastic theory of community food webs. VI. Heterogeneous alternatives to the cascade model. *Theor Popul Biol* 37:55–90
- Cohen JE (1994) Marine and continental food webs: three paradoxes? *Philos Trans Roy Soc* 343:57–69
- Cohen JE, Briand F (1984) Trophic links of community food webs. *Proc Natl Acad Sci USA* 81:4105–4109
- Cohen JE, Newman CM (1985) A stochastic theory of community food webs. 1. Models and aggregated data. *Proc R Soc Lond B* 224:421–448
- Cohen JE, Briand F, Newman CM (1990a) *Community food webs: data and theory*. Springer, Berlin Heidelberg New York
- Cohen JE, Luczak T, Newman CM, Zhou ZM (1990b) Stochastic structure and nonlinear dynamics of food webs: qualitative stability in a Lotka-Volterra cascade model. *Proc R Soc Lond B* 240:607–627
- Cohen JE, Pimm SL, Yodzis P, Saldaña J (1993) Body sizes of animal predators and animal prey in food webs. *J Anim Ecol* 62: 67–78

- Collett D (1991) Modelling binary data. Chapman & Hall, London
- Copas JB (1983) Plotting p against x . *Appl Stat* 32:25–31
- Duin RPW (1976) On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Trans Comp C* 25:1175–1179
- Elton C (1927) *Animal ecology*. MacMillan, New York
- Epanechnikov VA (1969) Nonparametric estimation of a multivariate probability density. *Theor Probab Appl* 14:153–158
- Everitt B (1977) *The analysis of contingency tables*. Chapman & Hall, London
- Fowler CW, MacMahon JA (1982) Selective extinction and speciation: their influence on the structure and functioning of communities and ecosystems. *Am Nat* 119:480–498
- Gallopini GC (1972) Structural properties of food webs. In: Patten BC (ed) *Systems analysis and simulation in ecology*, vol 2. Academic Press, New York, pp 241–282
- Gasser T, Müller HG, Mammitzsch V (1985) Kernels for nonparametric curve estimation. *J R Stat Soc B* 47:238–252
- Gittleman JL (1985) Carnivore body size: ecological and taxonomic correlates. *Oecologia* 67:540–544
- Habbema JDF, Hermans J, van der Broek K (1974) A stepwise discrimination program using density estimation. In: Bruckman G (ed) *Compstat 1974*. Physica, Vienna, pp 100–110
- Hall SJ, Raffaelli D (1991) Food web patterns: lessons from a species-rich web. *J Anim Ecol* 60:823–842
- Hall SJ, Raffaelli D (1993) Food webs: theory and reality. *Adv Ecol Res* 24:187–239
- Hansen B, Bjornsen PK, Hansen PJ (1994) The size ratio between planktonic predators and their prey. *Limnol Oceanog* 39:395–403
- Härdle W (1990) *Applied nonparametric regression*. Cambridge University Press, Cambridge, UK
- Hastie TJ, Tibshirani RJ (1990) *Generalized additive models*. Chapman & Hall, London
- Hutchinson GE (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *Am Nat* 93:145–159
- Jonsson T (1998) Food webs and the distribution of body sizes. *Linköping Studies in Science and Technology*. Dissertation No 535. Linköping, Sweden
- Jonsson T, Ebenman (1998) Effects of predator-prey body size ratios on the stability of food chains. *J Theor Biol* 193:407–419
- Lawton JH (1989) Food webs. In: Cherrett JM (ed) *Ecological concepts: the contribution of ecology to an understanding of the natural world*. Blackwell, Oxford
- Lawton JH, Warren PH (1988) Static and dynamic explanations for patterns in food webs. *Trends Ecol Evol* 3:242–245
- MacDonald N (1979) Simple aspects of food web complexity. *J Theor Biol* 80:577–588
- Martinez ND (1993) Effects of resolution on food web structure. *Oikos* 66:403–412
- Martinez ND, Lawton JH (1995) Scale and food-web structure—from local to global. *Oikos* 73:148–154
- May RM (1973) *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, NJ
- Mendenhall W, Scheaffer RL, Wackerly DD (1986) *Mathematical statistics with applications*. Duxbury Press, Boston, Mass
- Nadaraya EA (1964) On estimating regression. *Theory Prob Appl* 10:186–190
- Paine RT (1988) On food webs: road maps of interactions or grist for theoretical development? *Ecology* 69:1648–1654
- Pearson K (1904) *On the theory of contingency and its relation to association and normal correlation*. *Drapers' Co Memoirs, Biometric Ser No 1*, London
- Peters RH (1983) *The ecological implications of body size*. Cambridge University Press, Cambridge, UK
- Pimm SL (1982) *Food webs*. Chapman & Hall, London
- Pimm SL, Lawton JH (1977) Number of trophic levels in ecological communities. *Nature* 268:329–331
- Pimm SL, Lawton JH, Cohen JE (1991) Food web patterns and their consequences. *Nature* 350:669–674
- Polis GA (1991) Complex trophic interactions in deserts: an empirical critique of food web theory. *Am Nat* 138:123–155
- Polis GA (1994) Food webs, trophic cascades and community structure. *Aust J Ecol* 19:121–136
- Polis GA, Winemiller KO (1996) Food webs: what do they tell us about the world? In: Polis GA, Winemiller KO (eds) *Food webs: integration of patterns and dynamics*. Chapman & Hall, New York
- Sheldon RW, Prakash A, Sutcliffe WH (1972) The size distribution of particles in the ocean. *Limnol Oceanog* 17:327–340
- Sheldon RW, Sutcliffe WH, Prakash A (1973) The production of particles in the surface waters of the ocean with particular reference to the Sargasso Sea. *Limnol Oceanog* 18:719–733
- Silverman BW (1986) *Density estimation for statistics and data analysis*. Chapman & Hall, London
- Solow AR (1996) On the goodness of fit of the cascade model. *Ecology* 77:1294–1297
- Solow AR, Beet AR (1998) On lumping species in food webs. *Ecology* 79:2013–2018
- Stone M (1974) Cross-validated choice and assessment of statistical predictions. *J R Stat Soc B* 36:111–147
- Sugihara G (1982) Niche hierarchy: structure, organization, and assembly in natural communities. PhD thesis, Department of Biology, Princeton University
- Sugihara G (1984) Graph theory, homology and food webs. *Proc Symp Appl Math* 30:83–101
- Thiebaut ML, Dickie LM (1993) Structure of the body size spectrum of the biomass in aquatic ecosystems: a consequence of the allometry in predator-prey interactions. *Can J Fish Aquat Sci* 50:1308–1317
- Tukey JW (1961) Curves as parameters and touch estimation. *Proc 4th Berkeley Symp*, pp 681–694
- Vézina AF (1985) Empirical relationships between predator and prey size among terrestrial vertebrate predators. *Oecologia* 67:555–565
- Warren PH, Lawton JH (1987) Invertebrate predator-prey body size relationships: an explanation for upper triangular food webs and patterns in food web structure? *Oecologia* 74:231–235
- Watson GS (1964) Smooth regression analysis. *Sankhyá A26*:101–116
- Yodzis P (1984) Energy flow and the vertical structure of real ecosystems. *Oecologia* 74:231–235
- Yodzis P (1989) Patterns in food webs. *Trends Ecol Evol* 4:49–50
- Yodzis P, Innes S (1992) Body size and consumer-resource dynamics. *Am Nat* 139:1151–1175