

A Stochastic Theory of Community Food Webs. VI. Heterogeneous Alternatives to the Cascade Model

JOEL E. COHEN

*The Rockefeller University, 1230 York Avenue, New York,
New York 10021-6399*

Received June 16, 1989

This paper considers 14 models as possible explanations of three major structural trends observed in 113 community food webs: (1) The average proportion of species that are top species is roughly constant in webs with different total numbers of species. The same holds for the average proportions of intermediate and basal species. (2) The average proportion of links that are intermediate-top links is roughly constant in webs with different total numbers of species. The same constancy of proportions in webs with different numbers of species holds for the three other kinds of links. (3) The frequency distributions of chain lengths are unimodal with modes usually between two and five links. One model, the cascade model, explains these trends and other structural trends. The other 13 models relax in various ways the cascade model's assumption that the probability of predation is the same between every pair of potential predator and potential prey and its assumption that the average probability of predation is inversely proportional to the total number of species in the web. Numerical comparisons of the observed structural trends with the model predictions show that the only viable alternative to the cascade model is the predator-dominant, constant-column-sum, linear-link-scaling model. This model differs from the cascade model in assuming that the expected number of prey species of any potential predator is the same as that for any other potential predator. The biological lesson suggested by these findings is that potential predators appear to have a mean number of prey species that is closer to constant than to linearly increasing with rank in the cascade model. Models that assume that predation probabilities are determined primarily by the prey species or by the distance between species in a cascade make predictions that are inferior descriptions of the major structural trends. Models that assume the mean number of trophic links increases as the 1.35 power of the total number of species make predictions that are inferior to those of models that assume the mean number of trophic links increases linearly with the total number of species. © 1990

Academic Press, Inc.

1. THE CASCADE MODEL AND THE NEED FOR HETEROGENEOUS ALTERNATIVES

When a mathematical model succeeds in describing a variety of observed regularities, it becomes of interest to examine how sensitive the model's predictions are to changes in the model's assumptions. A systematic

investigation of the changes in predictions that follow from changes in assumptions is a sensitivity analysis.

A stochastic model that describes varied properties of community food webs is called the cascade model (Cohen, Briand and Newman, 1989). This paper reports a sensitivity analysis of two major assumptions of the cascade model. Thirteen alternatives to the cascade model are analyzed. These models result from a systematic $3 \times 2 \times 2 + 1$ design of combinations of specific hypotheses.

A sensitivity analysis, though a purely theoretical or mathematical activity, can usefully guide empirical observations. If a specific assumption of an original model can be replaced by a broad range of alternative assumptions without greatly affecting the model's predictions, then the model's ability to describe observed regularities provides little evidence in favor of the specific assumption as opposed to the alternatives, because these other assumptions could serve just as well. Moreover, an empirical study that produces evidence against the original assumption does not necessarily reject the class of models obtained by replacing it with one of the alternatives. On the other hand, if even minor variations of an original assumption substantially change the model's predictions, then the model's ability to describe observed regularities provides indirect evidence in favor of the original assumption compared to the minor variations; empirical

TABLE I

Observed Fractions of Trophic Species That Are Top, Intermediate, and Basal, and Observed Fractions of Trophic Links of Each Kind, in 113 Community Food Webs

113 community food webs	Observed number	Observed fraction	Cascade model predicted fraction
Basal species	353	0.186	0.231
Intermediate species	1038	0.546	0.537
Top species	511	0.269	0.231
all species	1902	1	1
Basal-intermediate links	1029	0.272	0.264
Basal-top links	230	0.061	0.114
Intermediate-intermediate links	1194	0.316	0.359
Intermediate-top links	1327	0.351	0.264
all links	3780	1	1

Note. A comparable tabulation for 62 webs has appeared before (I, p. 424). The tabulation for 113 webs has not appeared previously. Source of webs: Cohen, Briand, and Newman (1989). The predictions of the cascade model as $S \rightarrow \infty$ take the ratio of links to species as exactly 2 though the actual ratio is 1.99. The predicted fractions of the different kinds of species give the ratio of the expected numbers of proper basal, intermediate, and proper top species as a fraction of the expected number of nonisolated species.

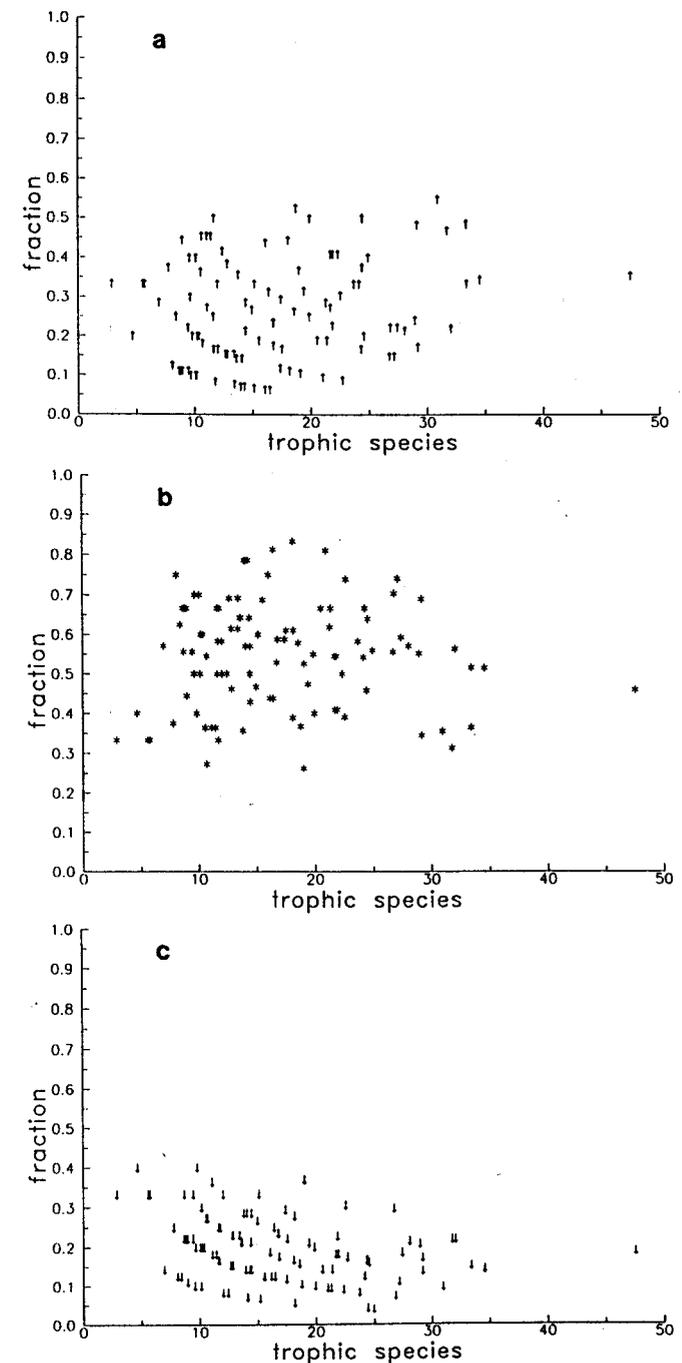


FIG. 1. Observed fractions of all trophic species that are (a) top, (b) intermediate, and (c) basal trophic species as a function of the total number of trophic species, in 113 community food webs. A comparable graph for 62 webs appeared in Briand and Cohen (1984). The graph for 113 webs has not appeared previously. Source of webs: Cohen, Briand, and Newman (1989).

evidence against the original assumption provides a substantial case against the model. Thus a sensitivity analysis sharpens the inferences that can be drawn from a model and can be useful in planning further empirical studies.

The sensitivity analysis to be reported here shows that some (but not all) relatively small variations in the assumptions of the cascade model change its predictions substantially. The analysis therefore provides indirect support for some (but not all) of the cascade model's original assumptions.

1.1. The Cascade Model

Why is it worth doing a sensitivity analysis of the cascade model? In 113 community food webs (hereafter "community food webs" will be abbreviated simply to "webs"), the cascade model describes qualitatively and quantitatively the fractions of all nonisolated trophic species that are basal, intermediate and top trophic species (Table I; compare Fig. 1 and Fig. 4a; see Appendix Section 7.2 for the definitions of nonisolated, basal,

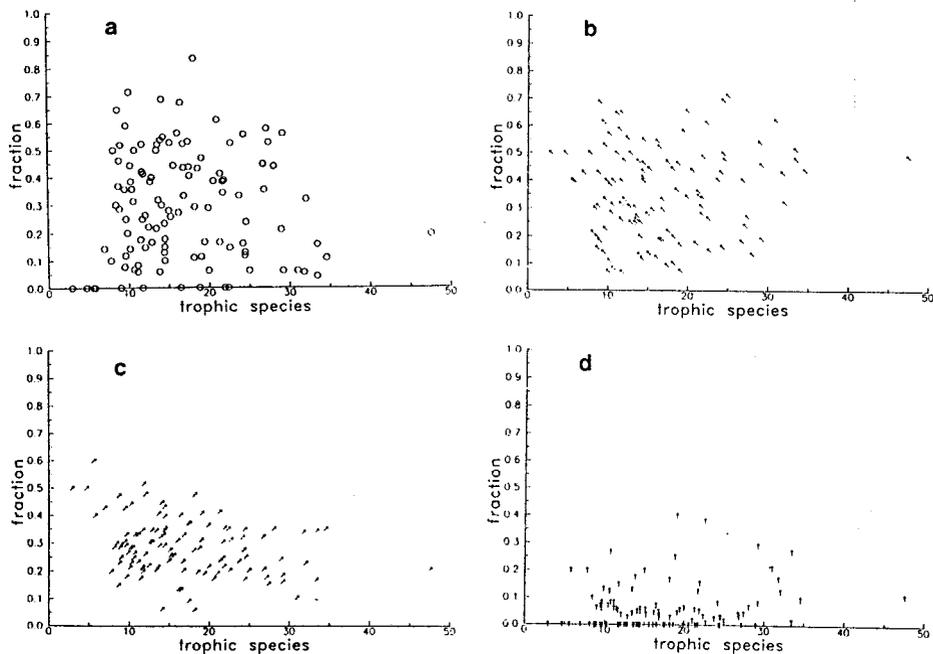


FIG. 2. Observed fractions of all trophic links that are (a) intermediate-intermediate, (b) intermediate-top, (c) basal-intermediate, and (d) basal-top trophic links as a function of the total number of trophic species, in 113 community food webs. A comparable graph for 62 webs appeared in Cohen and Briand (1984). The graph for 113 webs has not appeared previously. Source of webs: Cohen, Briand, and Newman (1989).

intermediate, and top trophic species). It describes the fractions of all trophic links that are basal-intermediate, basal-top, intermediate-intermediate, and intermediate-top trophic links (Table I; compare Fig. 2 and Fig. 5a). It describes the numbers of food chains of each length (compare Fig. 3 and Fig. 6a). It explains Hutchinson's (1959) observation that food chains are typically short compared to the number of species in a web. (Table II confirms that in all but seven of 113 webs, the modal chain length fell in the range from two to five links.) The cascade model accounts for the

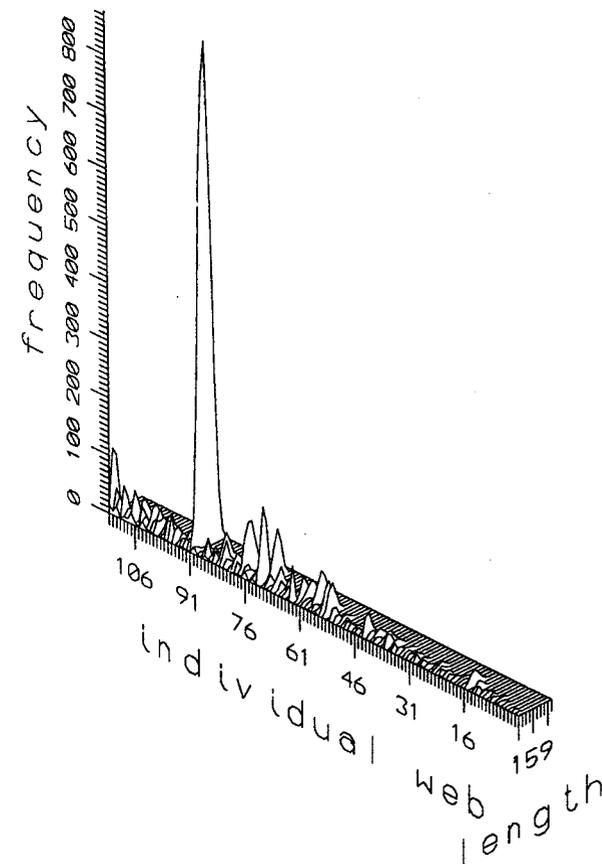


FIG. 3. Observed numbers of chains with one to nine links in 113 community food webs. No chains longer than nine links were observed. Each slice parallel to the axis labeled "length" is the observed frequency distribution for one web. The height of a slice at each value of length is the observed number of chains of that length. The webs are ranked from the smallest (web 1) to the largest (web 113) number of species. The frequency distributions were recalculated for this paper from the corrected predation matrices in Cohen, Briand, and Newman (1989).

observed declines in the frequencies of interval and triangulated niche overlap graphs, from near one for very small webs to near zero for webs of more than 35 species. These claims are documented in the previous papers of this series (Cohen and Newman, 1985, hereafter referred to as I; Cohen *et al.*, 1985, hereafter referred to as II; Cohen *et al.*, 1986, hereafter referred to as III; Newman and Cohen, 1986, hereafter referred to as IV; Cohen and Palka, 1989, hereafter referred to as V; all of which are included in Cohen, Briand, and Newman, 1989).

What does the cascade model assume? It assumes that the trophic species (hereafter simply "species") in a community are ordered in a cascade or hierarchy. Any species can consume only those species below it and can be consumed only by those species above it in the cascade. Operationally, if there are S species in the web, the cascade model assumes a labeling of the species from 1 to S in such a way that whenever a species labeled i is eaten by a species labeled j , then $i < j$. This assumption excludes the possibility of trophic cycles, e.g., cases where species i eats species i or where i eats j and j eats i . Moreover, the cascade model assumes that for any two species i and j with $i < j$, the probability that j eats i is p . The probability p is a positive constant that is the same for every pair of species i and j . Whether j actually eats i is statistically independent of all other eating relations in the web. When webs with different total numbers S of species are compared, the cascade model assumes that p depends inversely on S according to $p = c/S$. The positive constant c is the same for every S .

TABLE II
Frequency Distribution, over 113 Community Food Webs,
of the Modal Chain Length

Modal chain length	Number of webs
1	4
2	44
2.5	3
3	35
3.5	4
4	16
5	4
6	3

Note. Each web's modal chain length is the length of chain that occurred with largest frequency in that web. When two or more lengths occurred with frequency equal to the maximal frequency, the corresponding lengths were averaged; hence the three webs with modal chain length 2.5 and the four webs with modal chain length 3.5.

On the basis of 113 webs (Table I), the numerical value of c is currently estimated as $2 \times \text{links/species} = 2 \times 3780/1902 = 3.97$, very close to 4. In all the numerical computations in this paper, the value of c is taken as exactly 4.

In summary, the cascade model assumes (i) ordering: the prior existence of a labeling or cascade of species that limits the possible feeding relations; (ii) equiprobability: a constant probability of a link between any two species for which a link is possible; (iii) independence: the existence of a link for any pair of species is independent of the existence of a link for any other pair of species; and (iv) reciprocal scaling: the probability of a link from any species to another species with a higher number in the ordering depends on the number of species S in the web according to c/S .

1.2. The Need for Alternatives

A case can be made against each of the cascade model's four assumptions.

(i) The cascade model excludes cycles because cycles were rare and cannibalism was inconsistently reported in the 113 webs analyzed and because, when cycles were permitted in several alternatives to the cascade model, the predicted fractions of top and basal species were far from those observed (I). Nevertheless, contrary to the assumed acyclic ordering, cannibalism is widespread in nature. Many but not all cases of cannibalism consist of larger or later stages in the life cycle of a biological species eating smaller or earlier stages. In such cases, the different stages should be distinguished as different trophic species, even though they belong to a single biological species. Ultimately, decomposers cause all food chains to be part of large cycles, though decomposers and the cycles they produce were largely neglected in the webs described by the cascade model. Because it omits such cycles, the cascade model must be interpreted as relevant only to the cycle-free portion of webs.

(ii) The assumption of a constant probability of a trophic link from any species i to any species j for which $i < j$ in the cascade means that the expected number of predators or consumers of species i is $(S-i)p = (S-i)c/S$ and the expected number of prey of species i is $(i-1)p = (i-1)c/S$, where c is currently estimated as 4. These numbers range from 0 to nearly c . It may or may not be plausible to assume this range in the predatory and defensive abilities of consumer and resource species. Schoener (1989) argues that there is an upper bound on the number of species which consumer species can consume, and that prey species are limited in the number of consumer species they can defend against. The former upper bound is consistent with the cascade model's

upper bound on the mean number of prey; the latter limit is not consistent with the cascade model's upper bound on the mean number of consumers.

(iii) The assumption of independence between trophic links means that, in the model, all the species might choose to eat the same single prey species, or even might choose to eat nothing at all. Such biologically implausible events are unlikely according to the model, however.

A hypothetical example of direct evidence against independence would be a finding that the predation matrices of real webs with a given fixed number of species showed a negative correlation between the presence of a link from species 4, say, to species 5, and the presence of a link from species 3 to species 4. Such a correlation would show a lack of independence between different links: Studies of real food webs at this level of detail have not been done yet.

(iv) The assumption of reciprocal scaling, $p = c/S$, appears to be justified directly by empirical observations that connectance, which is directly proportional to p , falls in proportion to $1/S$ when webs with increasing numbers S of species are compared (Rejmánek and Starý, 1979; Pimm, 1982; Auerbach, 1984). It is trivial to show that the assumption of reciprocal scaling implies, and is implied by, the linear link-species scaling "law," which states that the expected number of trophic links in a web is directly proportional to the number of trophic species in the web (Cohen and Briand, 1984). However, it is easier empirically to detect deviations from a straight line than from a hyperbolic curve, so the linear link-species scaling law is a better curve to examine for discrepancies than the hyperbolic curve for connectance. On the basis of 40 webs, Briand (1983) estimated that the expected number of links was proportional to $S^{1.1}$; on the basis of 113 webs, fitting a straight line by least squares to the logarithms of the numbers of links and species, Cohen, Briand, and Newman (1986) estimated that the expected number of links was proportional to $S^{1.36}$. In general, if the expected number of links $E(L)$ is proportional to $S^{1+\epsilon}$, where $\epsilon > 0$, then the probability p of a link (or the connectance) is proportional to $E(L)/S^2 = S^{\epsilon-1}$.

Alternative assumptions about cycles (hypothesis i) have been considered before (I). There is little direct evidence against independence (hypothesis iii). So the following sensitivity analysis will consider alternatives to the cascade model's assumptions (ii) and (iv), which are equiprobability and linear link-species scaling.

1.3. The Criteria for Evaluating Alternatives

Alternative models will be evaluated by comparing their predictions with the predictions of the cascade model in three areas where the cascade model seems well supported by data.

First, the predicted fractions of nonisolated species that are top, intermediate, and basal should approximate well the observed proportions of each kind of species (Table I) and should be insensitive to (that is, nearly independent of) the number of species in the web over the observed range of numbers of species (Fig. 1).

Second, the predicted fractions of all trophic links that are basal-intermediate, basal-top, intermediate-intermediate, and intermediate-top trophic links should approximate well the observed proportions of each kind of trophic link (Table I) and should be insensitive to the number of species in the web over the observed range of numbers of species (Fig. 2).

In these first two areas, alternative models will be evaluated in the range from 10 to 70 species by increments of five species. Except for a few probably incomplete webs with a very small number of reported species, this range covers the available data.

Third, the expected numbers of food chains of each length should approximate well a typical observed distribution of chain length for a web of typical size, around 17 trophic species (Fig. 3). (The mean number of trophic species per real web is 16.8.) The predicted frequency distribution of chain length should be unimodal, with the mode between two and five links with very few predicted chains of eight links or longer.

The cascade model makes other successful predictions (V), but these three areas suffice for an initial evaluation.

2. SPECIFIC HETEROGENEOUS MODELS

This section enumerates some alternatives to the cascade model that are biologically motivated and simple to analyze.

Let S denote the number of trophic species in a community, where $4 \leq S < \infty$. The probability that species j eats species i will be written p_{ij} , for all i and j from 1 to S . The matrix P with (i, j) -element p_{ij} is called the predation probability matrix. The cascade assumption is equivalent to the assumption that $p_{ij} = 0$ whenever $j \leq i$. Because the elements on the main diagonal and below it are zero, the predation probability matrix is said to be strictly upper triangular.

The cascade model assumes that, whenever $i < j$, $p_{ij} = p$. A general heterogeneous model, analyzed in the Appendix, assumes that whenever $i < j$, p_{ij} depends arbitrarily on i and j , provided, of course, that $0 \leq p_{ij} \leq 1$. The alternatives to the cascade model to be considered now are intermediate in orderliness between the homogeneous cascade model and the general heterogeneous model. These models are specified by three hierarchical conditions on the values of p_{ij} for $i < j$. A first condition (Section 2.1) specifies whether predators, prey, or distances in the cascade

between predators and prey determine the probabilities of predation. A second condition (Section 2.2) specifies, e.g., in a predator-dominant model, the expected number of prey of different predators. A third condition (Section 2.3) specifies how the expected number of trophic links grows with increasing numbers of species.

2.1. Predator Dominance, Prey Dominance, and Distance Dominance

We shall consider in turn the possibilities that predation probabilities are determined by the consumer species, or by the prey species, or by the distance in the ordering between the predator and the prey.

First, suppose that each consumer is equally likely to consume any of the prey species available to it, but because of different behavioral capacities and morphology, different predators have different probabilities of preying on the species available to them. Then the model is said to be predator-dominant:

$$\text{Predator dominance: } p_{ij} = b_j > 0, \quad \text{for } j = 2, \dots, S \text{ and } i < j. \quad (2.1.1)$$

Next, suppose that the chances of predation are determined by the relative abundances or defensive abilities of the available prey species, so that different prey species have different probabilities of being preyed on but, as a first approximation, each prey is equally like to be preyed on by any of its possible consumers. Then the model is said to be prey-dominant:

$$\text{Prey dominance: } p_{ij} = a_i > 0, \quad \text{for } j = 2, \dots, S \text{ and } i < j. \quad (2.1.2)$$

Predator-dominant models with $p_{ij} = b_j$ and prey-dominant models with $p_{ij} = a_i$ are both special cases of models with $p_{ij} = a_i b_j$. Such models might be called predator-prey-independence models, by analogy with similar models for two-way contingency tables. The analysis of predator-prey-independence models will be deferred until it is clear whether predator-dominant or prey-dominant models are capable of explaining the data.

Finally, suppose that the chance that species i is eaten by species j is determined by the difference between the position of i and j in the cascade of species. If position in the cascade correlates with body size, as has been suggested (Warren and Lawton, 1987; Cohen, 1989), it would not be surprising if the difference in body size between consumer and prey dominated the probability of predation. A distance-dominant model assumes that it is only the difference between the cascade positions of two species that determines the probability that one feeds on the other:

$$\text{Distance dominance: } p_{ij} = d_{j-i} > 0, \quad \text{for } j - i = 1, \dots, S - 1. \quad (2.1.3)$$

The probabilities that species j will prey on other species appear in column j of the predation probability matrix P . Predator-dominant models assume that within any column, all predation probabilities are the same. The probabilities that species i will be eaten by other species are contained in row i of P . Prey-dominant models assume that within any row, all predation probabilities are the same. The probabilities of feeding between pairs of species i and j with $i < j$ that differ by a fixed amount $d = j - i$ in their positions in the cascade appear in what is defined here as the d th diagonal of P . For example, the first diagonal lies just above the main diagonal of P , and the $(S - 1)$ st diagonal consists solely of the element p_{1S} . (This definition differs from the definition that is customary in some parts of matrix theory.) Distance-dominant models assume that within any diagonal, all predation probabilities are the same.

2.2. Constant Line Sums and Increasing Line Sums

Collectively, columns, rows, and diagonals will be referred to as lines. Define the length of a line as the number of nonzero elements in it. Because P is strictly upper triangular, column 1 and row S have length 0, column 2 and row $S - 1$ have length 1, and so on: column S and row 1 have length $S - 1$. The sums of the predation probabilities in the i th row, the i th diagonal, and the i th column may be interpreted, respectively, as the expected number of consumers of species i , the expected number of trophic links from any species to another species that is i positions higher in the cascade, and the expected number of prey of species i .

In the cascade model, because $p_{ij} = p$ whenever $i < j$, each of these sums, i.e., the sum of the probabilities in any line (the line sum), is p times the line length. So row i and diagonal i have line sum $p(S - i)$ and column i has line sum $p(i - 1)$.

One biologically plausible alternative is to suppose that row or column sums are constant, regardless of their length. In a predator-dominant model, for example, suppose that as a result of anatomical or behavioral adaptations every predator, whether high or low in the cascade, has an equal expected number of prey species. This hypothesis may be viewed as an interpretation of Pimm's (1982, p. 89) explanation for the hyperbolic decline of connectance: "Suppose each species in a community feeds on a number of species of prey that is independent of the total number of species in the community." In a prey-dominant model, constant row sums would occur if a consumed species minimized its expected number of predators but could not reduce the expected number below a level that is the same for all species. In a distance-dominant model, diagonal sums are constant if, the further away the prey from the predator in the cascade (i.e., the shorter the diagonal), the more likely the predator is to eat the prey.

Thomas W. Schoener (personal communication, 6 March 1989) suggested a justification for an increase in predation probability with distance in the cascade, based on optimal foraging theory (Schoener, 1969, 1971) and the hypothesis that the cascade ordering of species corresponds to an ordering by size: the further a potential prey species from a potential predator, the smaller the prey relative to the predator, and therefore the lower the prey's probability of avoiding predation (e.g., by defense or escape).

To achieve constant line sums, it is necessary that the elements in a line be inversely proportional to its length:

$$\text{Constant line sums: } p_{ij} = \text{constant}/(\text{line length}), \quad 1 \leq i < j \leq S. \quad (2.2.1)$$

(The rule (2.2.1) encounters a difficulty in short lines if the constant on the right side exceeds one, since probabilities may not exceed one. A minor fix is required, which is explained in the Appendix.) Constant line sums imply, for example, that the higher a predator is in the cascade, the lower that predator's probability is of consuming any particular prey species, because the more prey species there are below it. Thus the longer the column, the lower the probabilities in it.

It also seems biologically reasonable to consider the opposite association between line length and the probability of predation; e.g., the longer the column, the higher the probabilities in it. In a predator-dominant model, suppose that predators higher in the cascade have increasing probabilities of preying on every species below them (suppose they are more "voracious" because, for example, their body sizes are larger and demand more energy). In a prey-dominant model, suppose that the species at the bottom of the cascade have the highest probability per species of being preyed on by each possible predator (because, for example, the smaller the organism, the more numerous the individuals or the greater the biomass or, as above, the greater the difficulty of avoiding predation). In a distance-dominant model, suppose that the less the difference between potential predator and potential prey in the cascade, the more likely it is that a given consumer will eat a given prey. A justification, again based on optimal foraging theory and ordering by size, is that a closer potential prey, being larger, is energetically more profitable to the potential predator, other things being equal (such as the costs of pursuit, capture and handling and population sizes) (Thomas W. Schoener, personal communication, 6 March 1989). A specific form of this hypothesis that causes the predation probabilities to vary like those in (2.2.1) is:

$$\text{Increasing line sums: } p_{ij} = \frac{\text{constant}}{S + 2 - \text{line length}}, \quad 1 \leq i < j \leq S. \quad (2.2.2)$$

2.3. Linear and Superlinear Link Scaling

Link scaling refers to the way the expected number of trophic links in a web depends on the total number of species. The expected number of trophic links equals the sum of the elements in the predation probability matrix. The expected number of trophic links will be denoted by $E(L)$.

According to the cascade model, each positive element of P is $p = c/S$ and there are $S(S-1)/2$ such elements above the main diagonal. Therefore $E(L) = (c/S)[S(S-1)/2] = (c/2)(S-1)$. Since $c=4$, the expected number of links is twice the number of species (minus one). Linear link scaling specifies that this relation continues to hold even when the elements of P are heterogeneous:

$$\text{Linear link scaling: } \sum_{i,j} p_{ij} = (c/2)(S-1). \quad (2.3.1)$$

Fitting a power law of the form $E(L) = aS^b$ by nonlinear least squares to the data on links and species given by Briand and Cohen (1987) gives an alternative superlinear scaling rule:

$$\text{Superlinear link scaling: } \sum_{i,j} p_{ij} = 0.69S^{1.35}. \quad (2.3.2)$$

The data on links and species of Briand and Cohen (1987) do not clearly speak for either of (2.3.1) or (2.3.2). For small and moderate numbers of species, the two predicted curves are almost identical, so webs with large numbers of species are required to distinguish the alternatives, and unfortunately there are very few webs with large numbers of species. Accurately described webs with large numbers of species will resolve the choice in the future.

2.4. Assembling Models

A specific heterogeneous model may now be chosen to order, like a three-course meal, by picking one option for dominance (predator, prey, or distance), one option for line sums (increasing or constant), and one option for link scaling (linear or superlinear). In principle, 12 models are possible. In addition, superlinear link scaling may be combined with homogeneous predation probabilities (as in the cascade model) to give a superlinear homogeneous model. Thus 13 alternatives to the cascade model will be considered. The exact formulas for each model appear in the Appendix.

3. DUALITY OF PREDATOR-DOMINANT AND PREY-DOMINANT MODELS

Not all of the models that can in principle be composed from the preceding menu turn out to make independent predictions about the

proportions of species, the proportions of links, or chain lengths. There is a close connection between the predictions of predator-dominant models and the corresponding prey-dominant models.

Any web may be turned upside down by reversing the direction of all the feeding relations in it and renumbering the species from top to bottom. Renumbering from top to bottom means that the number i is replaced by $S+1-i$, so that the first shall be last and the last shall be first. When one web results from turning another upside down in this way, we say that one web is the dual of the other, or the two webs are dual to one another.

Two web models (in the class of heterogeneous cascade models being considered here) with predation probability matrices X and Y are said to be dual to one another if the probability x_{ij} of a link from species i to species j in the first model equals the probability $y_{S+1-j, S+1-i}$ of a link from species $S+1-j$ to species $S+1-i$ in the second model, for all species i and j . Less formally, two web models (or their predation probability matrices X and Y) are dual to one another if the ensemble of predation matrices generated according to X is statistically indistinguishable from the ensemble of webs generated according to Y after the latter the webs are turned upside down.

For example, if a predator-dominant model and a prey-dominant model have the same rule for line sums and the same rule for link scaling, the two models are dual to each other. Specifically, a predator-dominant model with constant column sums is dual to the prey-dominant model with constant row sums if both have linear link scaling. The summary statistics of any two dual models are closely related, according to results proved in the Appendix.

First, the expected number of top species in a predator-dominant model equals the expected number of basal species in the dual prey-dominant model, and vice versa. The same equality holds for the expected numbers of proper top and proper basal species (see Appendix Section 7.2 for the definitions). The expected numbers of isolated species are the same in two dual models, and therefore the expected numbers of nonisolated species are also the same in two dual models. It follows that, in displaying the predictions of heterogeneous alternatives to the cascade model, it suffices to display the expected fractions of all nonisolated species that are top and basal for only one member of a pair of dual models, since the corresponding fractions for the other member of the pair are the same after reversing top and basal.

Similarly, the expected number of links, the variance in the number of links, and the expected number of intermediate-intermediate links in a predator-dominant model are equal, respectively, to the same quantities in the dual prey-dominant model. The expected number of basal-intermediate links in a predator-dominant model equals the expected number of

intermediate-top links in the dual prey-dominant model. It follows that it suffices to display the expected fractions of all trophic links in different subcategories for only one member of a pair of dual models, since the corresponding fractions for the other model are the same after top and basal are reversed.

Perhaps surprisingly, any two dual models have identical expected numbers of food chains of any given length, that is, the expected number of chains of length one in a predator-dominant model is the same as the expected number of chains of length one in the dual prey-dominant model, and similarly for chains of any other length. It follows that the predicted distribution of chain lengths cannot be used to distinguish a predator-dominant model from its dual prey-dominant model.

The cascade model, the superlinear homogeneous model, and each distance-dominant model are dual to themselves. Therefore, in these models, the expected number of top species equals the expected number of basal species, the expected number of proper top species equals the expected number of proper basal species, and the expected number of basal-intermediate links equals the expected number of intermediate-top links.

4. NUMERICAL RESULTS

The results of analyzing the alternative heterogeneous cascade models will be presented graphically, rather than in mathematical formulas. The results describe the proportions of nonisolated species that are basal, intermediate, and top species, the proportions of the four kinds of links, and the frequencies of chains of different lengths.

Altogether, 14 models are under consideration: the cascade model, the superlinear homogeneous model, and 12 heterogeneous alternatives. Because each predator-dominant model is dual to a corresponding prey-dominant model, the four predator-dominant models will be omitted from the graphs. Thus 10 models remain for graphing.

4.1. *Fractions of Top, Intermediate, and Basal Trophic Species*

Figure 4 shows five predicted ratios, $E(I)/E(N)$, $E(T)/E(N)$, $E(T_p)/E(N)$, $E(B)/E(N)$, and $E(B_p)/E(N)$, according to 10 models, as function of the expected numbers of nonisolated species, which are derived from the numbers of species ranging from $S=10$ to $S=70$. Here $E(I)$ denotes the expected number of intermediate species, $E(T)$ denotes the expected number of top species, $E(T_p)$ denotes the expected number of proper top species, $E(B)$ denotes the expected number of basal species, $E(B_p)$ denotes the expected number of proper basal species, and $E(N)$ denotes the expected number of nonisolated species. Though the quantities plotted are

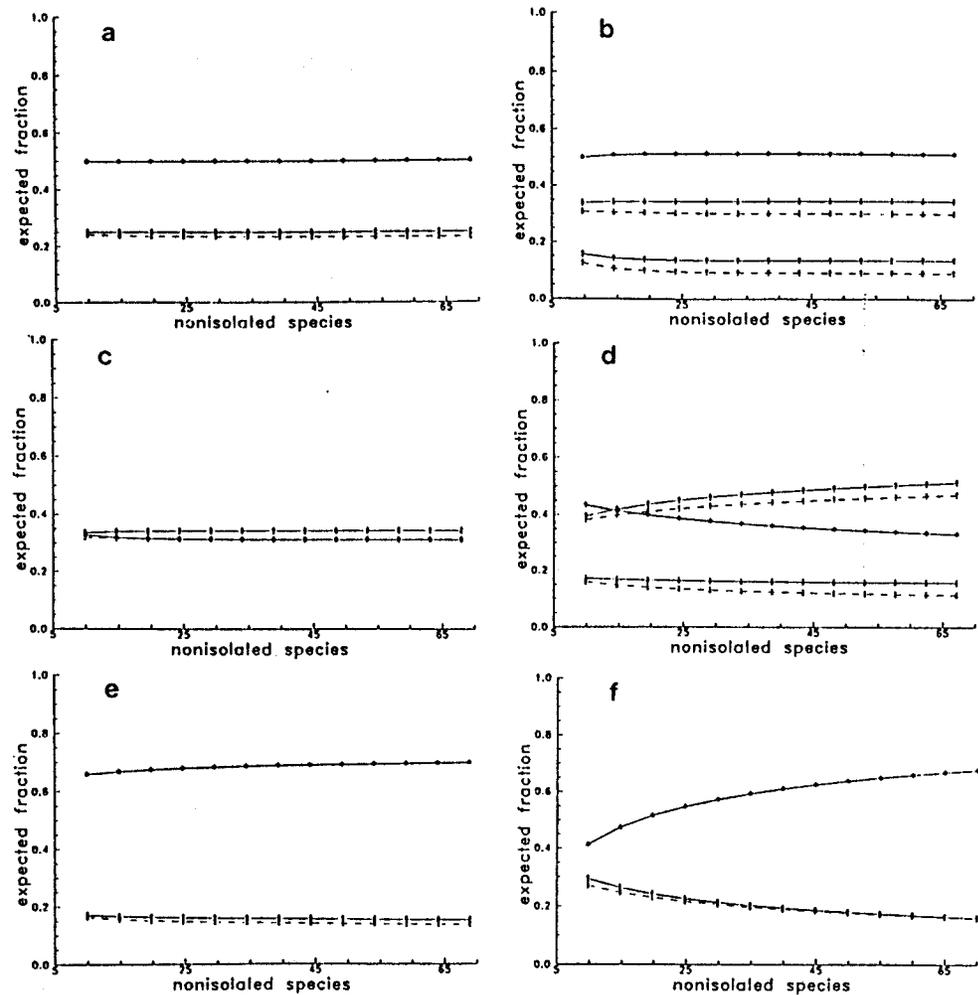


FIG. 4. Expected number of top, intermediate, basal, proper top, and proper basal species as a fraction of expected number of nonisolated species, for various numbers of expected nonisolated species. Computations were done for $S=10(5)70$. The abscissa plots the expected number of nonisolated species $E(N)$, not S . Key to symbols: solid line with asterisk is $E(I)/E(N)$; solid line with \uparrow is $E(T)/E(N)$; solid line with \downarrow is $E(B)/E(N)$; dashed line with \uparrow is $E(T_P)/E(N)$; dashed line with \downarrow is $E(B_P)/E(N)$. When one model is dual to another, only one of the two models is included in the plot; the results for the other model are identical except for the exchange of \uparrow and \downarrow . Models: (a) cascade model; (b) prey-dominant, constant-row-sums, linear-link-scaling; (c) distance-dominant, constant-diagonal-sums, linear-link-scaling; (d) prey-dominant, increasing-row-sums, linear-link-scaling; (e) distance-dominant, increasing-diagonal-sums, linear-link-scaling; (f) homogeneous superlinear-link-scaling; (g) prey-dominant, increasing-row-sums, superlinear-link-scaling; (h) distance-dominant, increasing-diagonal-sums, superlinear-link-scaling; (i) prey-dominant, constant-row-sums, superlinear-link-scaling; (j) distance-dominant, constant-diagonal-sums, superlinear-link-scaling.

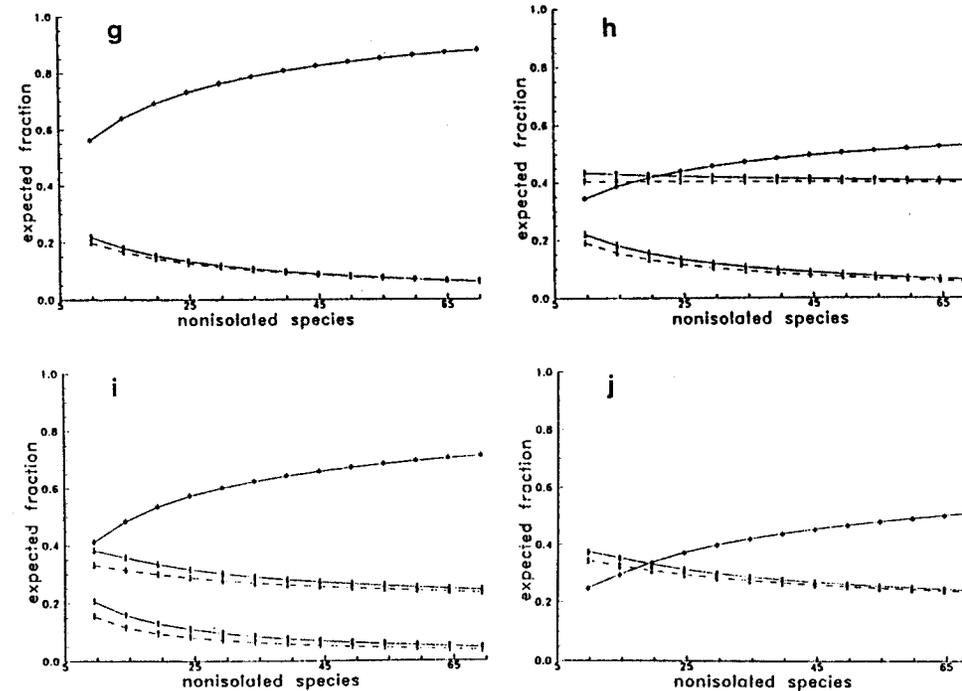


FIG. 4—Continued

ratios of expected values, it will be convenient to refer to them simply as proportions. Because of the numerical closeness between the proportions of top and proper top species, the discussion of results will not distinguish between them; similarly for basal and proper basal species.

The ratios of expected values will be compared against observations, each of which is a single ratio. For the crude qualitative purposes of this analysis, this comparison is reasonable in the light of Slutsky's theorem (e.g., Cramér, 1946, p. 255). Analytical calculation of the expected ratios, e.g., $E(I/N)$, seems very difficult, but numerical results could be obtained in particular cases, where needed, by simulation.

All five models with superlinear link scaling display clear increasing trends in the proportions of intermediate species. There is no compelling evidence for an increasing trend in the fraction of intermediate species in the data (Fig. 1). All five of these models also display decreasing trends, varying in strength, in the proportions of *both* top and basal species. No one has yet claimed to see in food web data a decreasing trend in the proportion of top species.

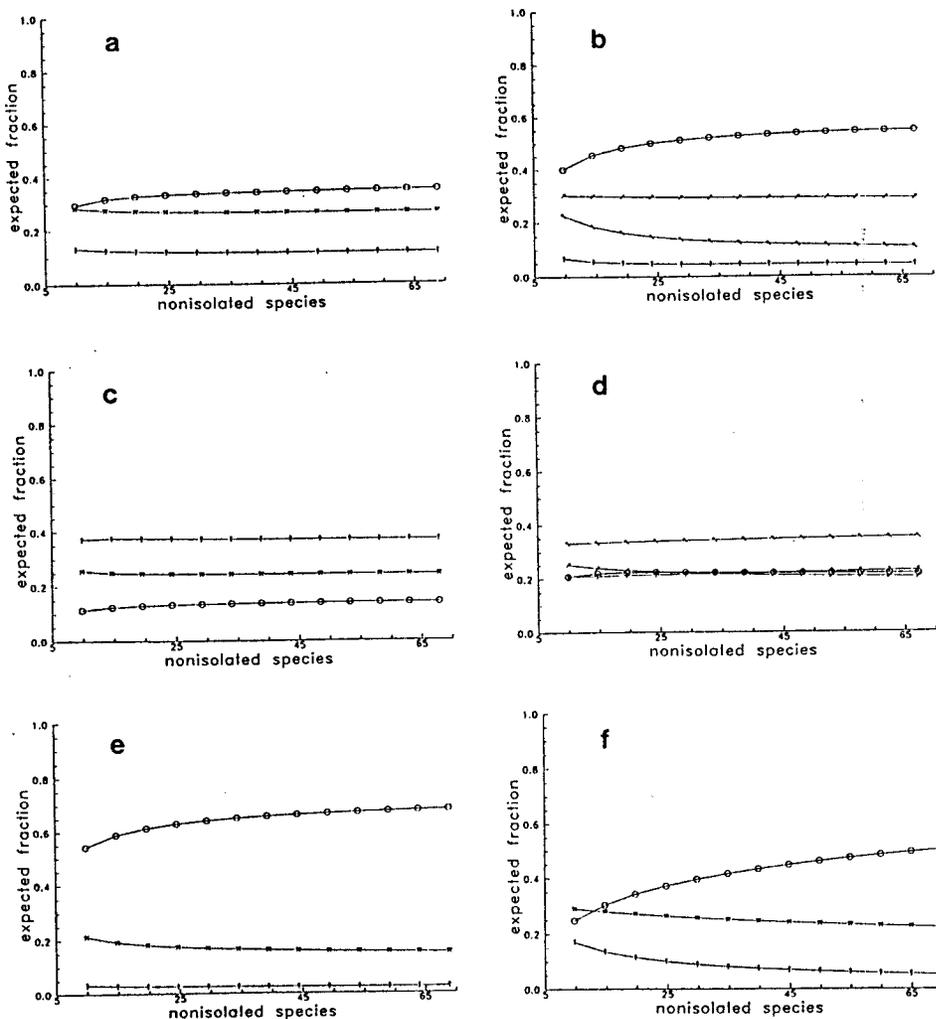


FIG. 5. Expected number of basal-intermediate, basal-top, intermediate-intermediate, and intermediate-top links as a fraction of expected number of links, for various numbers of expected nonisolated species. Computations were done for $S = 10(5)70$. The abscissa plots nonisolated species $E(N)$, not S . Key to symbols: 2 o'clock arrow is $E(L_{BI})/E(L)$; 10 o'clock arrow is $E(L_{IT})/E(L)$; \uparrow is $E(L_{BT})/E(L)$; open circle is $E(L_{II})/E(L)$. Plots for the dual models are identical to the plots shown except for the exchange of 2 o'clock arrows and 10 o'clock arrows. Models: (a) cascade model; (b) prey-dominant, constant-row-sums, linear-link-scaling; (c) distance-dominant, constant-diagonal-sums, linear-link-scaling; (d) prey-dominant, increasing-row-sums, linear-link-scaling; (e) distance-dominant, increasing-diagonal-sums, linear-link-scaling; (f) homogeneous superlinear-link-scaling; (g) prey-dominant, increasing-row-sums, superlinear-link-scaling; (h) distance-dominant, increasing-diagonal-sums, superlinear-link-scaling.

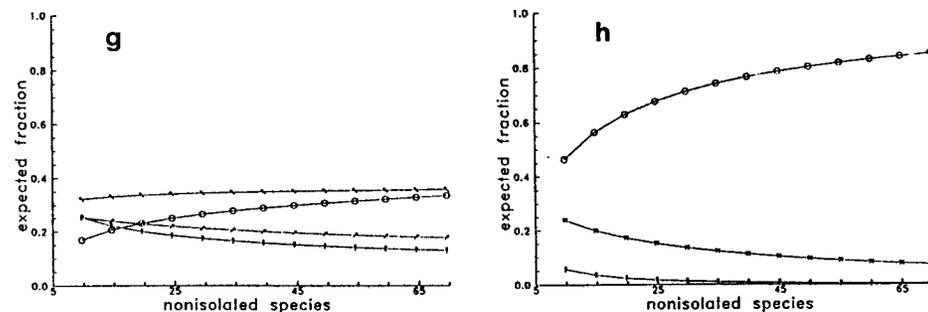


FIG. 5—Continued

The models with increasing line sums display less pronounced trends in the proportions of intermediate species with increasing nonisolated species, but the predicted proportions agree poorly with the data. Specifically, the prey-dominant, increasing-row-sum, linear-link-scaling model predicts a proportion of top species that equals or exceeds the proportion of intermediate species over most of the range of nonisolated species considered. The distance-dominant, increasing-row-sum, linear-link-scaling model predicts an excessive proportion of intermediate species and insufficient proportions of top and basal species.

The distance-dominant, constant-diagonal-sum, linear-link-scaling model predicts that the proportions of top, intermediate and basal species should be equal, contrary to observation. The prey-dominant, constant-row-sum, linear-link-scaling model predicts proportions of intermediate species that are close to those observed, but roughly twice as many basal species as top species, contrary to observation.

However, the dual model, namely, the predator-dominant, constant-column-sum, linear-link-scaling model predicts proportions of intermediate species that are close to those observed and more top species than basal species. The latter difference in the proportions of top and basal species is actually observed (Table I). The predictions of this model may easily be gleaned from Fig. 4b simply by switching the curves corresponding to top and basal species. This model predicts slightly more top species and slightly fewer basal than are actually observed but, compared to the predictions of the cascade model, the predictions of this model move in the direction of the data.

4.2. Fractions of Different Kinds of Trophic Links

Figure 5 shows four predicted ratios, $E(L_{BI})/E(L)$, $E(L_{IT})/E(L)$, $E(L_{BT})/E(L)$, and $E(L_{II})/E(L)$, according to eight models, for numbers of nonisolated species derived from numbers of species ranging from $S = 10$ to

$S = 70$. Here $E(L_{BI})$ denotes the expected number of basal–intermediate links, $E(L_{IT})$ denotes the expected number of intermediate–top links, $E(L_{BT})$ denotes the expected number of basal–top links, $E(L_{II})$ denotes the expected number of intermediate–intermediate links, and $E(L)$ denotes the expected number of all trophic links (the sum of the four preceding quantities). Though the quantities plotted are ratios of expected values, it will be convenient to refer to them simply as proportions. (Two superlinear models that are present in Fig. 4 are omitted from Fig. 5 as not requiring further analysis.)

Three models with superlinear link scaling display clear increasing trends in the proportions of intermediate–intermediate links. There is no compelling evidence for an increasing trend in the fraction of intermediate–intermediate links in the data (Fig. 2).

The models with increasing line sums display less pronounced trends in the proportion of intermediate–intermediate links with increasing non-isolated species, but the predicted proportions agree poorly with the data. Specifically, the prey-dominant, increasing-row-sum, linear-link-scaling model predicts proportions of intermediate–top links around 0.34 and equal proportions, around 0.22, of the other three kinds of links. This prediction is grossly contrary to the observed difference between the proportions of basal–top links and intermediate–intermediate links. The distance-dominant, increasing-diagonal-sum, linear-link-scaling model predicts an excessive proportion of intermediate–intermediate links.

The distance-dominant, constant-diagonal-sum, linear-link-scaling model predicts too few intermediate–intermediate links. The prey-dominant, constant-row-sum, linear-link-scaling model predicts too many intermediate–intermediate links and more basal–intermediate links than intermediate–top links, contrary to observation.

The dual model, namely, the predator-dominant, constant-column-sum, linear-link-scaling model is the alternative model that best approximated the observed proportions of species (reverse top and basal species in Fig. 5b). With regard to trophic links, this model predicts rising and excessively high proportions of intermediate–intermediate links and falling and excessively low proportions of basal–intermediate links. For both of these proportions, the cascade model's predictions are closer to the data than are the predictions of this model. However, reading from the graph, this model makes better predictions than does the cascade model of the proportions of basal–top links (predicted 0.06, observed 0.06, cascade model's predicted 0.11) and intermediate–top links (predicted 0.31, observed 0.35, cascade model's predicted 0.26). Thus, relative to the cascade model, the predictions of this model move in the direction of the data for basal–top and intermediate–top links, and away from the data for intermediate–intermediate links and basal–intermediate links.

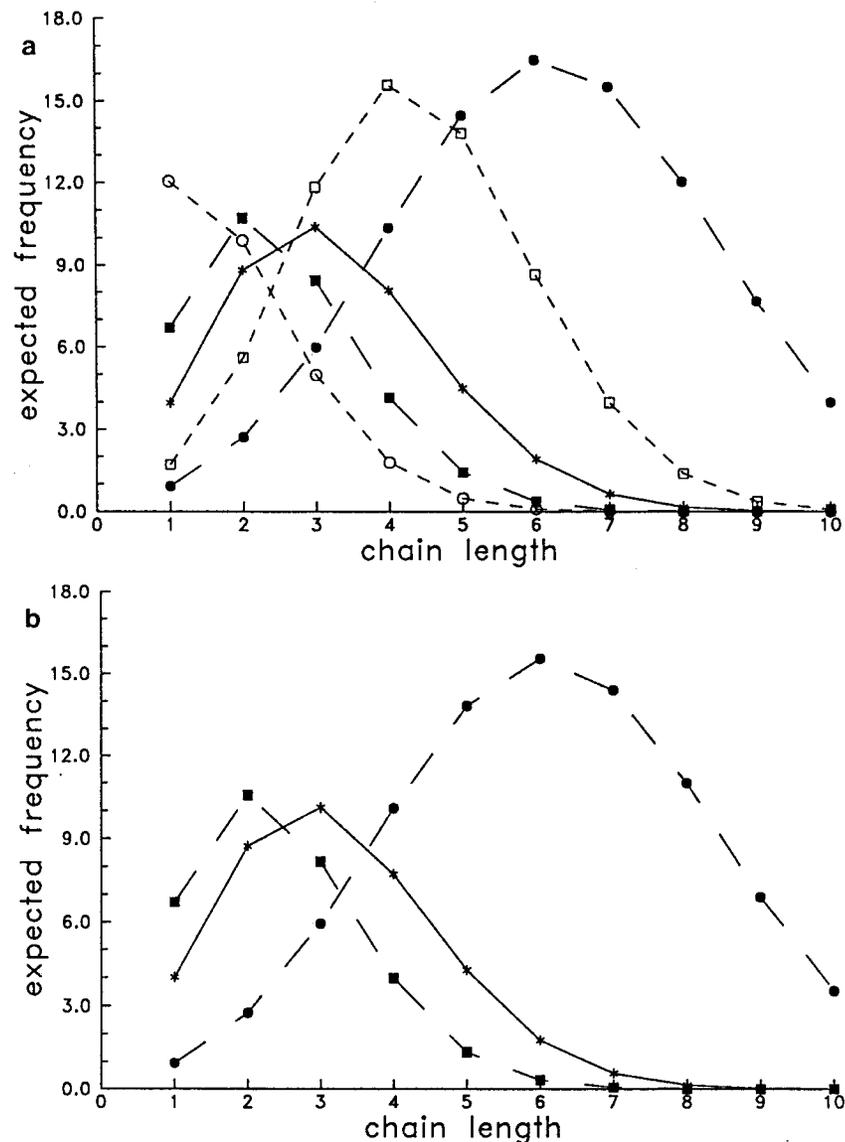


FIG. 6. Expected number of chains as a function of chain length in hypothetical food webs with 17 species. Upper figure shows linear-link-scaling models. The expected number of links, 32, is determined by the linear function $E(L) = 2(S - 1)$. Lower figure shows superlinear-link-scaling models. The expected number of links, 31.62, is determined by the power law $E(L) = 0.69S^{1.35}$. Key to symbols in both figures: solid line with asterisks is homogeneous (cascade) model; long dashed line with solid circle is distance-dominant increasing-diagonal-sum model and prey-dominant increasing-row-sum model. In upper figure only: short dashed line with open circle is distance-dominant constant-diagonal-sum model; long dashed line with open square is predator-dominant constant-column-sum model and prey-dominant constant-row-sum model.

4.3. Expected Numbers of Chains of Each Length

Figure 6 shows $E(C_n)$, the expected number of chains with n links for n from 1 to 10, according to eight models. In light of Table II those models where the modal chain length is less than two or greater than five may be rejected for general use. This relaxed standard of acceptability rejects all and only the distance-dominant models. Considering the observed variability in chain length distributions in different webs (Fig. 3), it seems prudent not to attempt to distinguish among the surviving models. However, predator-dominant and prey-dominant models with constant line sums have more long chains than the cascade model, while predator-dominant and prey-dominant models with increasing line sums have fewer long chains than either the cascade model or the homogeneous superlinear model. The chain length distributions of the cascade model and the homogeneous superlinear model are not identical, but are practically indistinguishable.

5. DISCUSSION

5.1. Baseline Data and Theory

This paper explores the range of alternative hypotheses which could explain the major structural trends observed in 113 community food webs. These trends are the proportions and scale-invariance of top, intermediate, and basal species (Fig. 1, Table I), the proportions and scale-invariance of the four kinds of links (Fig. 2, Table I), and the unimodal frequency distribution of chains lengths with mode between two and five links (Fig. 3, Table II). Individual webs deviate from these trends, sometimes substantially, but there are clear consistent patterns.

The interpretation that the data in Fig. 1 show scale-invariance in the proportions of top, intermediate, and basal species is not entirely uncontroversial. Schoener (1989) collected all but six of the webs listed by Briand and Cohen (1987), excluded a few that he considered deficient, edited the remaining 95 webs by different procedures, rather more freely, and interpreted his data as showing that the proportion of top species increased, and the proportion of basal species decreased, with increasing total species. Sugihara et al. (1989) analyzed 60 invertebrate-dominated webs. They found a marginally statistically significant decline in the proportion of basal species and a statistically non-significant rise in the proportion top species. Polis (1990) argued that with increasing sampling effort, the numbers of species will increase and the proportion of top species will go to zero, but this claim was not based on a comparison of existing webs. I do not see evidence of any such trends in Fig. 1. The resolution of these conflicting claims lies in the future. In the meantime, the simple summary

in the previous paragraph provides the empirical baseline for the following analysis.

The cascade model provides a simple qualitative and quantitative explanation of the structural trends in Figs. 1–3, Tables I and II, and others. A model of such simplicity cannot be expected to, does not aim to, and does not, describe precisely every individual web. Such a simple model may even fail to describe differences between biologically significant subgroupings of webs, if such exist. The goal of a simple model like the cascade model is to describe simply and quantitatively the major trends in webs.

5.2. Alternative Models

The cascade model is taken as a point of departure for alternative hypotheses. Thirteen variations relax in various ways two of the cascade model's assumptions, homogeneity and linear link scaling. Homogeneity means that the probability of a link, if not zero by the requirement of a rank ordering, is the same for every predator and prey. Linear link scaling means that the expected total number of links increases linearly with the number of species in the web.

Alternatives to homogeneity are the possibilities that predation probabilities are determined by each predator, or by each prey, or by the distance between predator and prey in the cascade. In the cascade model, a predator's probability of predation on an available prey is the same, regardless of the predator's position in the cascade. As alternatives, when predators determine the probabilities of predation, the probability of predation could decrease or increase as the predator's position in the cascade increases. Analogous changes in predation probabilities are considered for prey-dominant and distance-dominant models.

An alternative to linear link scaling is superlinear link scaling, meaning that the expected number of links in a web increases in proportion to the number of species raised to some power greater than one.

5.3. Results of Analysis

Three unanticipated findings emerged from examining all possible combinations of these alternative hypotheses. First, there is a kind of symmetry, here called duality, between predator-dominant and prey-dominant models which otherwise satisfy the same hypotheses. The proportion of nonisolated species that are top in a predator-dominant model equals the proportion of nonisolated species that are basal in the dual prey-dominant model, and vice versa. Because the cascade model, the homogeneous superlinear model, and the distance-dominant models are each dual to themselves, the proportions of nonisolated species that are top and that are basal are necessarily equal. Second, the expected number of chains of a given length

is the same for any two dual models. Hence the frequency distribution of chain length cannot be used to distinguish predator-dominant from prey-dominant models. Third, within the framework of heterogeneous cascade models, ad hoc adjustments of the predation probability matrix are required to reconcile the assumption that every predator has the same mean number of prey (or every prey has the same mean number of predators) with the assumption that the expected total number of links increases superlinearly with the number of species. These adjustments insure that the predicted probabilities of predation do not exceed one, and affect increasing numbers of predator-prey pairs as the total number of species increases.

5.4. Results of Computations

The alternative models are evaluated by the same standard as the cascade model: can they describe simply and quantitatively the major trends in webs? Numerical calculations (Figs. 4–6) show that among the 13 alternative models considered, only one is a potential competitor to the cascade model: the predator-dominant, constant-column-sum, linear-link-scaling model. This model predicts proportions of intermediate species that are close to those observed and, as observed, more top species than basal species. Compared to the predictions of the cascade model, the predicted species proportions of this model move in the direction of the data.

Pimm (1982) suggested that the observed excess of top over basal species is an artifact of ecologists' propensity to split predators and consumers into finer trophic categories than those they use for prey and resource species. However, the units of counting here are trophic species, and an observed excess of top over basal species persists. While the use of trophic species cannot completely eliminate the possibility of artifact, the predator-dominant, constant-column-sum, linear-link-scaling model provides the first alternative quantitative explanation of why more top than basal species are reported.

The predator-dominant, constant-column-sum, linear-link-scaling model also predicts proportions of basal-top and intermediate-top links that are closer to those observed than does the cascade model, and, unlike the cascade model, provides the first quantitative explanation of why more intermediate-top links are observed than basal-intermediate links. Unfortunately, this model is worse than the cascade model in predicting the proportions of intermediate-intermediate and basal-intermediate links. It also predicts a modal chain length of 4 links, which is perhaps further from the mode of the modes observed (Table II) than the modal chain length of 3 links predicted by the cascade model.

Only one model predicts a proportion of top species that rises and a proportion of basal species that falls (very slightly) with increasing total

species: the prey-dominant, increasing-row-sums, linear-link-scaling model. This same model predicts, as observed, more intermediate-top links than intermediate-intermediate links. In addition, this model predicts a modal chain length of 2 links, which is perhaps closer to the mode of the modes observed (Table II) than the modal chain length of 3 links predicted by the cascade model. Grossly contrary to observation, this model predicts fewer intermediate than top species in larger webs, and about equal proportions of intermediate-intermediate, basal-top, and basal-intermediate links. On balance, this model does not seem a serious competitor to the cascade model or to the predator-dominant, constant-column-sum, linear-link-scaling model.

5.5. Suggested Theoretical Studies

If the cascade model is better in some respects than the predator-dominant, constant-column-sum, linear-link-scaling model and in other respects the reverse is true, an obvious way to try to construct a model that combines features of both models is to take a weighted average of the predation probability matrices of each. For example, if X is the predation probability matrix of the cascade model and Y is the predation probability matrix of the predator-dominant, constant-column-sum, linear-link-scaling model, consider the predation probability matrix $aX + (1-a)Y$, where $0 < a < 1$. The most obvious candidate is $a = \frac{1}{2}$. The question that requires calculation is whether the model that results from this blending will combine the best or the worst features of each separate model or some of each.

A second class of models worth investigating contains the predator-prey-independence models mentioned in Section 2.1. Let us suppose, for a hypothetical example, that both the predator-dominant, constant-column-sum, linear-link-scaling model and the prey-dominant, increasing-row-sum, linear-link-scaling model have some empirical successes that make them attractive. In the first model, the nonzero elements in column j equal $2/(j-1)$ except for the first few columns (see (7.6.1)). In the second model, the nonzero elements in row i are proportional to $2/(i+1)$ (to see this, take the dual of (7.6.3)). It would be interesting to analyze a predator-prey-independence model with $p_{ij} = \text{constant}/[(j-1)(i+1)]$ or perhaps more simply $p_{ij} = \text{constant}/(ij)$, where the constant is chosen to give linear link scaling as a function of the total number of species.

To a theorist accustomed to deriving models from well-founded general principles, such as population genetics models based on Mendel's laws, the preceding suggestions may seem like crude empirical groping. Indeed they are. The reason is that at present there seem to be few links between the phenomenology of food web structure—the description of who eats what—and many of the well-established empirical generalizations of ecological physiology (the power-law scaling of metabolism as a function of

body size), biogeography (the scaling of species diversity with island area), age-structured population theory (asymptotic growth rates as a function of schedules of fertility and mortality), or species-abundance relations (the lognormal distribution). The integration of these diverse principles is a more ambitious goal for future theoretical work.

6. SUMMARY AND CONCLUSIONS

The purpose of this paper is to consider alternative possible explanations of three major structural trends observed in 113 community food webs: (1) The average proportion of species that are top species is roughly constant in webs with different total numbers of species. The same holds for the average proportions of intermediate and basal species. (2) The average proportion of links that are intermediate-top links is roughly constant in webs with different total numbers of species. The same constancy of proportions in webs with different numbers of species holds for the three other kinds of links. (3) The frequency distributions of chain lengths are unimodal with modes usually between two and five links. The data from all 113 webs in support of these trends are displayed here for the first time.

The cascade model provides simple qualitative and quantitative explanations of these and other structural trends. In a sensitivity analysis of the cascade model, two major assumptions are relaxed: the assumption that the probability of predation is the same between every pair of predator and prey for which predation is possible, and the assumption that the average probability of predation is inversely proportional to the total number of species in the web.

Analysis of models in which these assumptions are relaxed shows that the frequency distribution of chain length cannot be used to distinguish predator-dominant from prey-dominant models.

When the predictions of the cascade model and of 13 alternative models are compared numerically with the three major structural trends mentioned above, only two models remain viable: the cascade model and the predator-dominant, constant-column-sum, linear-link-scaling model.

The predator-dominant, constant-column-sum, linear-link-scaling model differs from the cascade model in assuming that the expected number of prey species of any potential predator is the same as that for any other potential predator. It predicts proportions of top, intermediate, and basal species that are closer to the data than the predictions of the cascade model. It also predicts proportions of basal-top and intermediate-top links that are closer to those observed than the predictions of the cascade model, and provides the first quantitative explanation of why more intermediate-top links are observed than basal-intermediate links. Unfortunately, this

model is worse than the cascade model in predicting the proportions of intermediate-intermediate and basal-intermediate links. It also predicts a modal chain length of 4 links, which is perhaps further from the mode of the modes observed than the modal chain length of 3 links predicted by the cascade model.

Thomas W. Schoener (personal communication, 6 March 1989) and Kenneth G. Schoenly (personal communication, 12 March 1989) remarked that the reported numbers of basal-intermediate and basal-top links may not be very reliable. Knowledge of basal species and detritivores is limited and standards for reporting these links are particularly inconsistent. Hence it may not be safe to discriminate among alternative models based on these characteristics. John Lawton (personal communication, 7 February 1989) emphasized the limited quality of all of the food web data, not merely the reported numbers of basal-intermediate and basal-top links. We share the critical skepticism of these, and many other, commentators. However, rather than give up any attempt to discriminate among models, we accept the data provisionally and draw conclusions that are subject to revision.

The biological lesson suggested by these findings is that predators appear to have a mean number of prey species that is closer to constant than to linearly increasing with position in a (hypothetical) cascade. Models that assume that predation probabilities are determined primarily by the prey species or by the distance between species in a cascade make predictions that are inferior descriptions of the major structural trends. Models that assume the mean number of trophic links increases as the 1.35 power of the total number of species make predictions that are inferior to models that assume the mean number of trophic links increases linearly with the total number of species.

7. APPENDIX: A HETEROGENEOUS CASCADE MODEL

7.1. A Heterogeneous Model

Let S be the number of species in a community, $4 \leq S < \infty$. The elements of the $S \times S$ predation matrix A are $a_{ij} = 1$ if species j eats i , $a_{ij} = 0$ if species j does not eat i . In modeling an ensemble of food webs, the elements a_{ij} of A are treated as random variables. The probability that $a_{ij} = 1$ is written $P\{a_{ij} = 1\}$.

The heterogeneous cascade model assumes that $P\{a_{ij} = 1\} = p_{ij}$, $0 \leq p_{ij} \leq 1$, where

$$p_{ij} = 0 \quad \text{if } i \geq j, \quad (7.1.1)$$

$$p_{ij} > 0 \quad \text{if } i < j, \quad (7.1.2)$$

and

the events $\{a_{ij} = 1\}$ are mutually independent for all $1 \leq i < j \leq S$. (7.1.3)

The matrix P with elements p_{ij} is called the predation probability matrix.

More general cascade models could weaken the assumption (7.1.3) of independence among trophic links or relax the strict inequality in (7.1.2) to $p_{ij} \geq 0$ for $i < j$. Such extensions will not be considered here.

Define the notation

$$q_{ij} = 1 - p_{ij} = P\{a_{ij} = 0\}, \quad \text{for } i, j = 1, \dots, S, \quad (7.1.4)$$

$$r_S = 1, \quad r_i = \prod_{j=i+1}^S q_{ij}, \quad \text{for } i = 1, \dots, S-1, \quad (7.1.5)$$

$$c_1 = 1, \quad c_j = \prod_{i=1}^{j-1} q_{ij}, \quad \text{for } j = 2, \dots, S. \quad (7.1.6)$$

The probability that row i of A is entirely 0 is given by r_i and the probability that column j of A is entirely 0 is given by c_j .

7.2. Categories of Species

A species is defined to be *not isolated* if it is joined by a trophic link to at least one other species, *top* if it is preyed on by no other species, *basal* if it consumes no other species, *intermediate* if it eats and is eaten by another species, *proper top* if it is top and not isolated, and *proper basal* if it is basal and not isolated. Let the number of each of these kinds of species in a web be denoted N , T , B , I , T_p , and B_p , respectively; these numbers are random variables. For $i = 1, \dots, S$, it is obvious that

$$P\{\text{species } i \text{ is not isolated}\} = 1 - r_i c_i,$$

$$P\{\text{species } i \text{ is top}\} = r_i,$$

$$P\{\text{species } i \text{ is basal}\} = c_i,$$

$$P\{\text{species } i \text{ is intermediate}\} = (1 - r_i)(1 - c_i) = 1 - r_i - c_i + r_i c_i,$$

$$P\{\text{species } i \text{ is proper top}\} = r_i(1 - c_i),$$

$$P\{\text{species } i \text{ is proper basal}\} = c_i(1 - r_i).$$

Consequently,

$$E(N) = S - \sum_{i=1}^S r_i c_i, \quad (7.2.1)$$

$$E(T) = \sum_{i=1}^S r_i, \quad \text{var}(T) = \sum_{i=1}^S r_i(1 - r_i), \quad (7.2.2)$$

$$E(B) = \sum_{i=1}^S c_i, \quad \text{var}(B) = \sum_{i=1}^S c_i(1 - c_i), \quad (7.2.3)$$

$$E(I) = 2S - E(T) - E(B) - E(N), \quad (7.2.4)$$

$$E(T_p) = E(T) + E(N) - S, \quad (7.2.5)$$

$$E(B_p) = E(B) + E(N) - S. \quad (7.2.6)$$

A group of species linked only to each other and not linked to another group of species in a web receives no special treatment here. Such isolated components of the web are extremely rare in theory and are absent from our data.

7.3. Categories of Links

Let L denote the total number of trophic links, i.e., the number of elements of A equal to 1, L_{BT} the number of basal-top links, L_{BI} the number of basal-intermediate links, L_{IT} the number of intermediate-top links, and L_{II} the number of intermediate-intermediate links. For $i < j$, there is a basal-top link from i to j if and only if there is a link from i to j (with probability p_{ij}) and j is top (with probability r_j) and i is basal (with probability c_i). By such arguments, it is obvious that

$$E(L) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}, \quad (7.3.1)$$

$$\text{var}(L) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij} q_{ij}, \quad (7.3.2)$$

$$E(L_{BT}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij} c_i r_j, \quad (7.3.3)$$

$$E(L_{BI}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij} c_i (1 - r_j), \quad (7.3.4)$$

$$E(L_{IT}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij} (1 - c_i) r_j, \quad (7.3.5)$$

$$E(L_{II}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij} (1 - c_i) (1 - r_j). \quad (7.3.6)$$

7.4. Chain Lengths

There is a (maximal) chain of length n , where n counts the number of links, that involves the $n+1$ species i_0, i_2, \dots, i_n if and only if $1 \leq i_0 < i_1 < \dots < i_n \leq S$; and there is a link from i_h to i_{h+1} for all

$h=0, \dots, n-1$; and i_0 is basal; and i_n is top. Let C_n denote the number of chains of length n . Then

$$E(C_n) = \sum_{i_0=1}^{S-n} \sum_{i_1=i_0+1}^{S-n+1} \cdots \sum_{i_h=i_{h-1}+1}^{S-n+h} \cdots \sum_{i_n=i_{n-1}+1}^S \prod_{h=0}^{n-1} p_{i_h i_{h+1}} c_{i_0} r_{i_n} \quad (7.4.1)$$

or

$$E(C_n) = \sum_{i_0=1}^{S-n} c_{i_0} \sum_{i_1=i_0+1}^{S-n+1} p_{i_0 i_1} \cdots \sum_{i_h=i_{h-1}+1}^{S-n+h} p_{i_{h-1} i_h} \cdots \sum_{i_n=i_{n-1}+1}^S p_{i_{n-1} i_n} r_{i_n}. \quad (7.4.2)$$

For different values of n , different numbers of summations are required to evaluate $E(C_n)$. A recursive function is a convenient way to program the computation.

7.5. Dual Predator-Dominant and Prey-Dominant Models

Let $X = (x_{ij})$ and $Y = (y_{ij})$ be two $S \times S$ predation probability matrices. Define X to be the dual of Y and write $X = \text{dual}(Y)$ if $x_{ij} = y_{S+1-j, S+1-i}$, for all $i, j = 1, 2, \dots, S$. The dual of Y is obtained by first transposing Y in the usual way and then changing the index of each row and each column from i to $S+1-i$, for $i = 1, \dots, S$. Alternatively, the dual of a matrix may be thought of as flipping the matrix over around the line from the $(n, 1)$ -element to the $(1, n)$ -element, just as the usual transpose of a matrix may be thought of as flipping the matrix over around the line from the $(1, 1)$ -element to the (n, n) -element. It is easy to check that $X = \text{dual}(Y)$ if and only if $Y = \text{dual}(X)$, and therefore that $X = \text{dual}(\text{dual}(X))$.

For example, if X is the predation probability matrix of a predator-dominant, constant-column-sum, linear-link-scaling model and Y is the predation probability matrix of a prey-dominant, constant-row-sum, linear-link-scaling model, then $X = \text{dual}(Y)$ and $Y = \text{dual}(X)$. The predation probability matrices of the cascade model and of the distance-dominant models are self-dual.

Let V stand for a random variable defined on a predation matrix, e.g., $V = N, T, B, I, T_P$, or B_P . Let $E(V|X)$ denote the expectation of V , given that X is the predation probability matrix.

THEOREM. *If $X = \text{dual}(Y)$, then*

$$E(N|X) = E(N|Y), \quad E(I|X) = E(I|Y), \quad (7.5.1)$$

$$E(T|X) = E(B|Y), \quad E(T_P|X) = E(B_P|Y), \quad (7.5.2)$$

$$E(L|X) = E(L|Y), \quad \text{var}(L|X) = \text{var}(L|Y), \quad (7.5.3)$$

$$E(L_{BT}|X) = E(L_{BT}|Y), \quad E(L_{II}|X) = E(L_{II}|Y), \quad (7.5.4)$$

$$E(L_{BI}|X) = E(L_{IT}|Y), \quad (7.5.5)$$

and

$$E(C_n|X) = E(C_n|Y), \quad n = 1, 2, \dots, S-1. \quad (7.5.6)$$

Proof. From the definition (7.1.5),

$$\begin{aligned} r_i(X) &= \prod_{j=i+1}^S (1-x_{ij}) = \prod_{j=i+1}^S (1-y_{S+1-j, S+1-i}) \\ &= \prod_{j=1}^{S-i} (1-y_{j, S+1-i}) = c_{S+1-i}(Y). \end{aligned} \quad (7.5.7)$$

Similarly,

$$c_j(X) = r_{S+1-j}(Y). \quad (7.5.8)$$

Therefore

$$E(T|X) = \sum_{i=1}^S r_i(X) = \sum_{i=1}^S c_i(Y) = E(B|Y). \quad (7.5.9)$$

The proofs of the remaining conclusions are too similar to bear repetition.

COROLLARY. *If $X = \text{dual}(X)$, then*

$$E(T|X) = E(B|X), \quad E(T_P|X) = E(B_P|X) \quad (7.5.10)$$

and

$$E(L_{BT}|X) = E(L_{IT}|X). \quad (7.5.11)$$

7.6. Exact Specifications of Alternative Models

This section specifies the predation probability matrices P of 13 alternatives to the cascade model. It is assumed throughout that $S \geq 6$ and that $p_{ij} = 0$ for $i \geq j$.

1. Predator-dominant, constant-column-sum, linear-link-scaling model.

The goal in specifying this model is to make each column (except the first, which is necessarily all zero) of the predation probability matrix have constant elements that add up to 2, so that the expected number of links is $2(S-1)$, as required by linear link scaling. However, column 2 has length one, so its sum cannot exceed one. To compensate, the first three elements of column 4 are filled with 1's. The remaining columns are filled with 2 divided by the length of the column, as desired. The effect of the

irregularity in columns 2-4 is negligible for large numbers of species. We have

$$p_{ij} = \begin{cases} 1 & \text{for } j=2, 3, 4 \quad \text{and} \quad i < j, \\ \frac{2}{j-1} & \text{for } j=5, \dots, S \quad \text{and} \quad i < j. \end{cases} \quad (7.6.1)$$

2. *Prey-dominant, constant-row-sum, linear-link-scaling model.* The predation probability matrix of this model is the dual of that in the previous model.

3. *Distance-dominant, constant-diagonal-sum, linear-link-scaling model.* The elements of the previous predation probability matrix are rearranged so that the elements in each diagonal are constant.

$$p_{ij} = \begin{cases} 1 & \text{for } j-i = S-3, S-2, S-1, \\ \frac{2}{S-j+i} & \text{for } j-i = 1, \dots, S-4. \end{cases} \quad (7.6.2)$$

4. *Predator-dominant, increasing-column-sum, linear-link-scaling model.* As an intermediate step, the strictly upper triangular matrix P' is calculated with elements

$$p'_{ij} = \frac{2}{S+2-j} \quad \text{for } j=2, \dots, S \quad \text{and} \quad i < j, \quad (7.6.3)$$

The elements of P' are then normalized to give linear link scaling:

$$P = \frac{2(S-1)P'}{\sum_{i,j=1}^S p'_{ij}}. \quad (7.6.4)$$

For large enough S , all the elements of P are probabilities. The proof is easy. According to Abramowitz and Stegun (1970, formula 6.1.3),

$$\gamma = \lim_{m \rightarrow \infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{m} - \log(m) \right), \quad (7.6.5)$$

where γ is Euler's constant, approximately 0.577. Therefore

$$\begin{aligned} \sum_{j=1}^S p'_{ij} &= 2 \left(\frac{1}{S+1-i} + \frac{1}{S-i} + \frac{1}{S-i-1} + \dots + \frac{1}{2} \right) \\ &\approx 2[\gamma - 1 + \log(S+1-i)], \end{aligned} \quad (7.6.6)$$

whence

$$\sum_{i,j=1}^S p'_{ij} \approx 2[(S-1)(\gamma-1) + \log(S!)]. \quad (7.6.7)$$

By Stirling's formula (Abramowitz and Stegun 1970, formula 6.1.38), the dominant term in $\log(S!)$ is $S \log(S)$, which grows faster than $2(S-1)$. Thus for large enough S , the maximal elements of P decline approximately in proportion to $1/\log(S)$. Hence they must eventually be probabilities.

If $S=6$, the sum on the left of (7.6.7) is 10.3 and the approximation on the right of (7.6.7) is 8.9306, so the approximation is not bad even for very small S . Numerical experimentation shows that $S=6$ is large enough to guarantee that the elements of P are all probabilities.

5. *Prey-dominant, increasing-row-sum, linear-link-scaling model.* The predation probability matrix of this model is the dual of that in the previous model.

6. *Distance-dominant, increasing-diagonal-sum, linear-link-scaling model.* The elements of the previous predation probability matrix are rearranged so that the elements in each diagonal are constant.

Superlinear-link-scaling models. If X is the predation probability matrix of one of the six models above or of the cascade model, a superlinear version of the model has predation probability matrix Y defined by

$$Y = X \frac{0.69S^{1.35}}{2(S-1)}. \quad (7.6.8)$$

This simple formula produces predation probability matrices with elements that do not exceed one for the three models with increasing line sums and for the superlinear homogeneous model, but not for the three models with constant line sums.

In the three models with *linear* link scaling and constant line sums, the largest element in the predation probability matrix is one, by construction, in the lines of length 1, 2, and 3 (see (7.6.1) and (7.6.2)). Therefore the largest element in the corresponding model with *superlinear* link scaling according to (7.6.8) is of order $S^{1.35-1}$, which increases beyond all bounds with increasing S . Numerical computation shows that, for S from 20 to 70, the maximal elements of the rescaled matrices exceed one but that the only elements that exceed one are those in lines of length 1, 2, and 3. Hence a further adjustment of (7.6.8) is required to make superlinear link scaling compatible with constant line sums.

7. *Predator-dominant, constant-column-sum, superlinear-link-scaling model.* Let Y be defined by (7.6.8), where X equals the predation proba-

bility matrix P given by (7.6.1). If Y has no elements greater than one, let Y be the predation probability matrix. Otherwise, let

$$\Delta = \sum_{x_{ij} > 1} (x_{ij} - 1)$$

be the total excess above one of the elements of X . For $20 \leq S \leq 70$, the elements in excess of one are those in columns 2, 3, and 4. So Δ must be evenly distributed over the remaining $S - 4$ columns numbered 5, 6, ..., S while the elements in columns 2, 3 and 4 are replaced by one. Since the length of column j is $j - 1$, each element in column j of Y is increased by $\Delta / [(S - 4)(j - 1)]$, $j = 5, 6, \dots, S$, while each element in columns 2, 3, and 4 of Y is replaced by one. The result is the predation probability matrix of the predator-dominant, constant-column-sum, superlinear-link-scaling model.

(It can be shown that for every fixed finite column number j , there exists a large enough species number S so that all elements in column j must exceed one when the rescaling (7.6.8) is applied to the predator-dominant, constant-column-sum, linear-link-scaling model. For large S , all columns j with $1 < j < 0.69S^{0.35} + 1$ have elements that exceed one. Asymptotically, for large S , more and more columns require ad hoc adjustment to avoid matrix elements that exceed one.)

8. *Prey-dominant, constant-row-sum, superlinear-link-scaling model.* The predation probability matrix of this model is the dual of that in the previous model.

9. *Distance-dominant, constant-diagonal-sum, superlinear-link-scaling model.* The elements of the previous predation probability matrix are rearranged so that the elements in each diagonal are constant.

10–13. *Increasing-line-sum, superlinear-link-scaling models and the superlinear homogeneous model.* Apply (7.6.8) to models 4, 5, 6, and the cascade model.

In the three models with increasing line sums, the largest element of P' before the normalization (7.6.4) is one, so the largest element of P is of order $1/\log(S)$, as explained above. Thus the largest elements in the corresponding matrices with superlinear link scaling are of order $S^{0.35}/\log(S)$, which increases beyond all bounds with increasing S , but increases very slowly. In the range of S from 10 to 70, the maximal elements of the predation probability matrices do not exceed one. The maximal elements decline from 0.63 when $S = 10$ to 0.52 when $S = 55$ and then increase slowly to 0.53 when $S = 70$. In this range, the hypotheses of superlinear link scaling and increasing line sums are consistent with each other and with dominance by predator, prey or distance. For sufficiently large S , however, superlinear link scaling and increasing line sums are not consistent.

Since the elements of the cascade model's predation probability matrix X are all $2/S$, the rescaled matrix Y of the superlinear homogeneous model has elements of the order of $0.69S^{1.35-2}$, which are probabilities that decrease with increasing S .

ACKNOWLEDGMENTS

I thank Zeng Zheng at the Southwestern Agricultural University, Beipei, Chongqing, Sichuan, People's Republic of China, for the questions he raised during my visit there in July, 1988; Thomas W. Schoener for prepublication drafts of his 1989 paper and for extremely detailed and helpful criticisms of previous drafts of my paper; and John H. Lawton and Kenneth G. Schoenly for comments on a previous draft. This work was supported by U.S. National Science Foundation Grant BSR 87-05047 and the hospitality of Mr. and Mrs. William T. Golden.

REFERENCES

- ABRAMOWITZ, M., AND STEGUN, I. A. 1970. "Handbook of Mathematical Functions," Dover, New York (7th Dover reprinting, with corrections).
- AUERBACH, M. J. 1984. Stability, probability and the topology of food webs, in "Ecological Communities: Conceptual Issues and the Evidence" (Donald R. Strong, Jr., Daniel Simberloff, Lawrence G. Abele, and Anne B. Thistle, Eds.), pp. 412–436, Princeton Univ. Press, Princeton, NJ.
- BRIAND, F., AND COHEN, J. E. 1984. Community food webs have scale-invariant structure, *Nature* **307**, 264–266.
- BRIAND, F., AND COHEN, J. E. 1987. Environmental correlates of food chain length, *Science* **238**, 956–960 and **243**, 239–240.
- COHEN, J. E. 1978. "Food Webs and Niche Space," Princeton Univ. Press, Princeton, NJ.
- COHEN, J. E. 1989. Food webs and community structure, in "Perspectives on Ecological Theory," (Jonathan Roughgarden, Robert M. May, Simon Levin, Eds.), pp. 181–202, Princeton Univ. Press, Princeton, NJ.
- COHEN, J. E., AND BRIAND, F. 1984. Trophic links of community food webs, *Proc. Nat. Acad. Sci. U. S. A.* **81**, 4105–4109.
- COHEN, J. E., BRIAND, F., AND NEWMAN, C. M. 1986. A stochastic theory of community food webs. III. Predicted and observed lengths of food chains. *Proc. R. Soc. London Ser. B* **228**, 317–353.
- COHEN, J. E., BRIAND, F., AND NEWMAN, C. M. 1989. "Community Food Webs: Data and Theory," *Biomathematics*, Vol. 20, Springer-Verlag, Berlin/New York.
- COHEN, J. E., AND NEWMAN, C. M. 1985. A stochastic theory of community food webs. I. Models and aggregated data, *Proc. R. Soc. London Ser. B* **224**, 421–448.
- COHEN, J. E., AND NEWMAN, C. M. 1988. Dynamic basis of food web organization, *Ecology* **69**, 1655–1664.
- COHEN, J. E., NEWMAN, C. M., AND BRIAND, F. 1985. A stochastic theory of community food webs. II. Individual webs, *Proc. R. Soc. London Ser. B* **224**, 449–461.
- COHEN, J. E., AND PALKA, Z. J. 1989. A stochastic theory of community food webs. V. Interspecificity and triangulation in the trophic niche overlap graph. *Amer. Nat.*
- CRAMÉR, H. 1946. "Mathematical Methods of Statistics," Princeton Univ. Press, Princeton, NJ.

- NEWMAN, C. M., AND COHEN, J. E. 1986. A stochastic theory of community food webs. IV. Theory of food chain lengths in large webs, *Proc. R. Soc. London Ser. B* **228**, 355–377.
- PIMM, S. L. 1982. "Food Webs," Chapman & Hall, London.
- POLIS, G. A. 1990. Complex trophic interactions in deserts: An empirical assessment of food web theory, *Amer. Nat.*
- REJMÁNEK, M., AND STARY, P. 1979. Connectance in real biotic communities and critical values for stability of model ecosystems, *Nature* **280**, 311–313.
- SCHOENER, T. W. 1969. Models of optimal size for solitary predators, *Amer. Nat.* **103**, 277–313.
- SCHOENER, T. W. 1971. Theory of feeding strategies. *Ann. Rev. Ecol. Syst.* **2**, 369–404.
- SCHOENER, T. W. 1989. Food webs from the small to the large, *Ecology* **70**, 1559–1589.
- SUGIHARA, G., SCHOENLY, K., AND TROMBLA, A. 1989. Scale invariance in food web properties. *Science* **245**, 48–52.
- WARREN, P. H., AND LAWTON, J. H. 1987. Invertebrate predator–prey body size relationships: An explanation for upper triangular food webs and patterns in food web structure? *Oecologia* **74**, 231–235.