

VARIATION OF TROPHIC FRACTIONS AND CONNECTANCE IN FOOD WEBS

PAUL A. MURTAUGH AND JEFFREY P. KOLLATH¹

Department of Statistics, Oregon State University, Corvallis, Oregon 97331 USA

Abstract. We present and illustrate methods of analyzing food web structure based on generalized linear models. Applied to the published food webs of Schoenly et al., these methods reveal strong associations of trophic fractions and connectance with the number of species in the web. Furthermore, the nature of these relationships depends on the habitat type of the web. Measures of goodness-of-fit of the statistical models reveal pronounced overdispersion, i.e., web-to-web variation in trophic fractions and connectance that greatly exceeds the variation attributable to the number of species and habitat type of the web. These results suggest it may be difficult to identify laws governing food web properties over a wide range of organisms and habitats.

Key words: *connectance; food webs; generalized linear models; logistic regression; multinomial regression; scale invariance.*

INTRODUCTION

A food web is a set of organisms in a community and the trophic interactions among them. Webs consist of top species (having prey but no predators), intermediate species (having both predators and prey), and basal species (having predators but no prey). “Connectance” is the number of trophic links among species in the web divided by the number of possible links. Food webs have been depicted in terms of both biological species (Schoenly et al. 1991, Havens 1992) and “trophic species,” i.e., groups of organisms having identical predators and prey (Briand and Cohen 1984).

Several quantitative properties of food webs have been touted as “laws” that apply to webs from diverse ecological communities (e.g., see Cohen et al. 1990), although other workers have doubted the quality of the data upon which these generalizations are based (Polis 1994, Winemiller and Polis 1996). Prominent among the generalizations are the claims that the fractions of species in the different trophic levels (Briand and Cohen 1984) and the connectance (Martinez 1992) do not vary with the number of species in the web, i.e., they are “scale-invariant.”

In order to make valid generalizations about food webs, ecologists need to agree on (1) which data sets are of sufficiently high quality to reflect food web properties accurately; and (2) what quantitative and/or statistical methods are appropriate for exploring those properties. The care taken by Schoenly et al. (1991) in assembling and considering possible biases inherent in their food webs, and the attention focused on this data

set by a variety of workers (e.g., Sugihara et al. 1989, Bengtsson 1994, Martinez 1994), suggest that this is an important set of webs to consider. In exploring the structure of this and other sets of food webs, some workers have used little or no statistical analysis (Sugihara et al. 1989, Havens 1992), while others have used linear regression and other statistical tools (Martinez 1992, Bengtsson 1994). Different quantitative approaches can lead to quite different interpretations of the same data (e.g., compare Havens 1992, Martinez 1993a, and Murtaugh 1994), so it would be useful to have some agreement on what the best methods of analysis are.

Here we apply methods based on generalized linear models to the data of Schoenly et al. (1991). These methods sidestep some of the problems of ordinary linear regression applied to fractional data, and they allow for easy modeling of the associations of covariates such as habitat type with patterns in food webs. We test the hypotheses that trophic fractions and connectance do not vary with the number of species in the web, we explore the variation of food web properties among habitat types, and we compare our conclusions to those of others applying different methodologies to the same data set.

METHODS

The data

The data considered here come from 95 insect-dominated food webs described by Schoenly et al. (1991). We restricted our attention to the subset of 61 community webs, i.e., webs consisting of all of the trophically connected species in a community. (We refer to the nodes of the webs as “species,” even though some of them represent coarser aggregations than biological species.) Some of the 61 webs come from the same

Manuscript received 28 March 1996; revised 30 September 1996; accepted 30 September 1996.

¹ Present address: Department of Preventive and Societal Medicine, University of Nebraska Medical Center, Omaha, Nebraska 68198-4350 USA.

TABLE 1. Summary of information on the 16 community webs from Schoenly et al. (1991).

Number	Habitat category	Number of webs	Range of S
1	Phytotelmata	15	5-19
2	Plant-herbivore-predator	9	7-25
3	Gall-gallmaker-parasitoid	4	36-62
4	Rotting wood-saprophage-predator	4	36-90
5	Carrion-necrovore-predator	13	3-50
6	Dung-coprovore-predator	7	16-35
7	Aquatic	9	11-60

systems monitored at different times, potentially violating the assumption of independence in statistical analysis. Working with a derived set of 43 independent web summaries, Bengtsson (1994) found patterns similar to those obtained with the full set of 61 webs, so we will work with the full data set here.

The major reasons for focusing on the data of Schoenly et al. (1991) are the care taken by the authors in selecting webs for inclusion, and their consideration of possible biases arising from uneven taxonomic resolution and sampling effort. In addition, the classification of the webs into seven habitat categories (Table 1) allows us to explore whether the relationships between food web properties and the number of species in the web differ across habitats. Finally, the attention focused on this data set by other workers, and the discrepancies in their conclusions about food web properties, invite the application of what we think are better methods of analysis.

The data set discussed by Sugihara et al. (1989) consists of 60 webs from the Schoenly et al. (1991) data, with minor differences in the specifications of some of the webs. Although Martinez (1994) suggested that these differences might influence assessments of scale invariance, we found that our models of trophic fractions and connectance were very similar between data sets.

Trophic fractions

To analyze the change in trophic fractions with number of species S in the web, we used a form of multinomial regression (Murtaugh 1994). If N is the number of webs in the data set and n_{ij} is the number of species in the j^{th} level of the i^{th} web ($i = 1, \dots, N; j = 1, 2, 3$), we can think of the observed numbers of species in each trophic level as a trinomial random vector, with probability density function

$$f(n_{i1}, n_{i2}, n_{i3} | S_i) = \frac{S_i!}{n_{i1}!n_{i2}!n_{i3}!} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} p_{i3}^{n_{i3}}, \quad (1)$$

where p_{ij} is the probability of a species in web i ending up in trophic level j . Any particular arrangement of n_{i1} , n_{i2} and n_{i3} species among the three trophic levels occurs

with probability $p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} p_{i3}^{n_{i3}}$ and there are $S_i!/(n_{i1}!n_{i2}!n_{i3}!)$ such arrangements.

To model the possible dependence of the cell probabilities on S_i and other covariates $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ik})'$, we can write

$$\begin{aligned} \text{logit } p_{i1} &\equiv \log\left(\frac{p_{i1}}{1 - p_{i1}}\right) = \alpha_1 + \beta_1 S_i + \boldsymbol{\gamma}'_1 \mathbf{X}_i \\ \text{logit } p_{i2} &= \alpha_2 + \beta_2 S_i + \boldsymbol{\gamma}'_2 \mathbf{X}_i, \end{aligned} \quad (2)$$

where $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are $k \times 1$ vectors of regression coefficients. (The third probability is then obtained by difference: $p_{i3} = 1 - p_{i1} - p_{i2}$.) By including covariates that are indicators for the different habitat types, and terms for the interactions of those covariates with S , we can test for the effects of habitat type on the relationships between the trophic fractions and S .

The details of statistical inference with this model, which is an example of a generalized linear model, are given in McCullagh and Nelder (1989). The method of maximum likelihood can be used to obtain parameter estimates, their standard errors, and a measure of residual variation called the deviance. All programming and calculations were done in the S language (Becker et al. 1988).

With certain qualifications, the deviance for an appropriate model fit to a data set should have a χ^2 distribution with the number of degrees of freedom associated with the residuals ($2N - 2[k + 2]$) for the model given by Eqs. 1 and 2). Thus, the deviance should, on average, be approximately equal to its degrees of freedom. In all of the models presented below, the deviance greatly exceeds its degrees of freedom, suggesting overdispersion; i.e., an apparent tendency for the multinomial probabilities in Eqs. 1 and 2 to vary among webs more than would be expected from the differences in covariates alone. Consequently, a dispersion parameter estimated as the Pearson statistic divided by the degrees of freedom was used to correct for the effects of overdispersion in all significance testing (see McCullagh and Nelder 1989:174).

A stepwise selection procedure was used to identify habitat groups that were statistically distinct with respect to their relationships between trophic fractions and S . First, a model was fit with just the covariate S . Then, for each of the seven habitat types, an indicator variable for the habitat and a term for its interaction with S were added, and the corresponding drop in deviance was calculated. The habitat variables leading to the largest drop in deviance were then added to the model, and the testing procedure was repeated for the remaining habitat types. This procedure continued until none of the remaining habitat variables produced a statistically significant improvement in the fit of the model.

Connectance

Connectance is defined as the number of links L in a food web divided by the maximum possible number

of links, given by S^2 if we include cannibalism (e.g., see Martinez 1992). The constant connectance hypothesis states that this ratio is equal to a constant c , which implies

$$\log L = \log c + 2 \log S. \quad (3)$$

Thus, a linear regression of $\log L$ against $\log S$ should have a slope of approximately 2 if the hypothesis is correct.

Generalized linear models can also be used to explore the relationship between L and S . If p_i is the probability that a potential link in web i is "engaged"—i.e., a trophic interaction is observed—then we can model this probability as a function of S_i and other covariates, \mathbf{X}_i , that are specific to the web:

$$\text{logit } p_i \equiv \log\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta S_i + \boldsymbol{\gamma}'\mathbf{X}_i. \quad (4)$$

If, for consistency with Schoenly et al. (1991), we ignore cannibalism and suppose that the maximum possible number of links is $S(S - 1)/2$, then logistic regression of the fractions of engaged links, $L_i/[S_i(S_i - 1)/2]$, against S_i and the other covariates provides estimates of the parameters in Eq. 4, their standard errors, and the deviance (Hosmer and Lemeshow 1989, McCullagh and Nelder 1989). The constant connectance hypothesis can then be assessed by testing whether the regression coefficient for S is equal to zero.

As with the trophic fractions described in the preceding section, overdispersion was common in the link data, and statistical testing was adjusted by means of a dispersion parameter estimated as the Pearson statistic divided by the degrees of freedom (see McCullagh and Nelder 1989:126). It should be noted that this adjustment inflates the variance estimates compared to those that are routinely reported by logistic regression packages, so that a particular association is less likely to be declared statistically significant. As described in the section on trophic fractions, stepwise variable selection was used to identify habitat groups having statistically distinct relationships of connectance to S .

RESULTS

Trophic fractions

Table 2 shows the results from applying the multinomial regression model (Eqs. 1 and 2) to the data of Schoenly et al. (1991), ignoring habitat type. Individually, the top and intermediate fractions show no significant change with the number of species S , but there is a statistically significant decrease in the basal fraction as S increases. A likelihood ratio test of the overall association of S with the trophic fractions is statistically significant ($P = 0.011$). The deviance from the model in Table 2 is 418, with 118 df, indicating substantial overdispersion.

When covariates for habitat type were added to the model in stepwise fashion, phytotelmata and carrion

TABLE 2. Regression equations from application of the multinomial model to the data from Schoenly et al. (1991), ignoring habitat type.

Level	Regression equation	<i>P</i> value for slope coefficient
Top	$\text{logit } \hat{p} = -0.318 + 0.00542S$	0.21
Intermediate	$\text{logit } \hat{p} = -0.378 + 0.00122S$	0.77
Basal	$\text{logit } \hat{p} = -1.365 - 0.02189S$	0.008

Note: The equations for the intermediate and basal levels are from a model based explicitly on those two fractions (i.e., having p_{i1} and p_{i2} in Eq. 2 corresponding to the intermediate and basal levels), while the equation for the top level is from a model based explicitly on the top and intermediate fractions. The overall P value for testing for any effect of S on the trophic fractions is 0.011 (likelihood ratio test).

habitats were judged to have web patterns that were statistically distinct from those of the other five habitats. Figs. 1–3 show the data and fitted regression lines for the three habitat groups and three trophic levels. One habitat group shows a significant decrease and another a significant increase in the top fraction with increasing S (Fig. 1); two groups show significant increases in the intermediate fraction with S (Fig. 2); and none of the three groups individually shows a significant trend in the basal fraction (Fig. 3), although collectively there is a highly significant decrease in this fraction with increasing S (cf. Table 2).

The deviance from the model distinguishing the three habitat groups is 288, with 110 df. This reflects a reduction in the overdispersion found in the model lacking habitat covariates, but still suggests that there are important differences in the multinomial probabilities among webs, above and beyond what can be explained by the varying numbers of species and types of habitat represented by the webs.

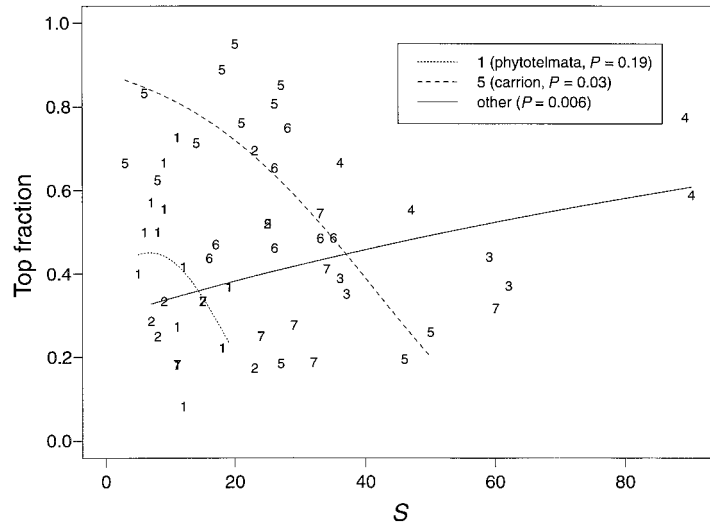
Connectance

The logistic regression model (Eq. 4) applied to the data without reference to habitat type indicates a highly significant decrease in the probability of a link being engaged with increasing S : $\text{logit } \hat{p} = -0.6072 - 0.0299S$ ($P < 0.0001$). The discrepancy between the deviance (461) and the residual degrees of freedom ($N - 2 = 59$) is enormous, indicating extreme overdispersion.

Stepwise variable selection led to a model identifying four habitat groups having distinct relationships of connectance to S : plant-herbivore-predator, rotting wood, aquatic webs, and other webs. Fig. 4 is a scatterplot of connectance vs. S , along with the fitted regression lines for the four habitat groups. Three of the four groups show statistically significant decreases in connectance with increasing S .

This full model still reflects substantial overdispersion (deviance of 250, with 53 df). Apparently, there

FIG. 1. Fraction of species in the top trophic level vs. number of species in the web. Plotted numbers indicate habitat type (see Table 1). Fitted lines for the three habitat groups judged to be distinct were calculated as $1 - p_1 - p_B$, where p_1 and p_B are the probabilities for the intermediate and basal levels, shown in Figs. 2 and 3. P values are for the hypothesis of no change in the top fraction with S . The regression coefficient for S in the combined data, ignoring habitat, is 0.0054 ($P = 0.21$).



are factors other than the number of species in the web and habitat type that cause connectance to vary considerably among webs.

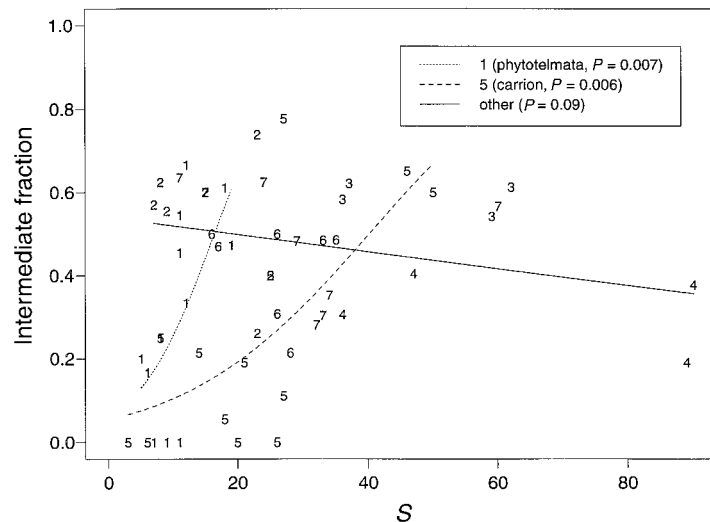
DISCUSSION

The use of multinomial regression for analyzing trophic fractions was discussed by Murtaugh (1994), but logistic regression, while increasingly used in other ecological applications (Trexler and Travis 1993), has not been applied to food web connectance, as far as we know. These methods, which are examples of generalized linear models, have several advantages over ordinary linear regression in the food web context: the response variables are properly modeled as whole numbers of species or links, rather than continuously varying quantities; the variance is modeled as a function of the mean, instead of as a constant; all predicted probabilities are between 0 and 1, which is not nec-

essarily true when ordinary linear regression is used; and overdispersion can be estimated and compensated for in significance testing. For these reasons, we believe that patterns identified with our approach that were not apparent in earlier analyses warrant the attention of ecologists.

There is clear evidence in this data set that trophic fractions vary with the number of species in the web, as judged from the global test for any change in the three fractions with S ($P = 0.011$) and the statistically significant decline in the basal fraction with increasing S ($P = 0.008$; Table 2, Figs. 1–3). Changes in one or more of the trophic fractions with S were also noted by Schoenly et al. (1991) and Martinez (1994). However, Sugihara et al. (1989), studying 60 of the webs included in the present data set, claimed the data were consistent with scale invariance of trophic fractions. Like some earlier workers, Sugihara et al. (1989) avoid-

FIG. 2. Fraction of species in the intermediate trophic level vs. number of species in the web. Plotted numbers indicate habitat type (see Table 1). Fitted lines for the three habitat groups were calculated from the multinomial regression model applied separately to each group. P values are for the hypothesis of no change in the intermediate fraction with S . The regression coefficient for S in the combined data, ignoring habitat, is 0.0012 ($P = 0.77$).



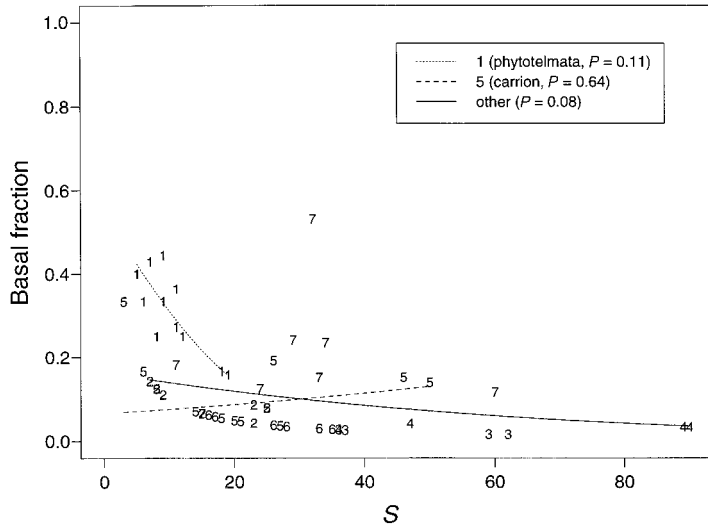


FIG. 3. Fraction of species in the basal trophic level vs. number of species in the web. Plotted numbers indicate habitat type (see Table 1). Fitted lines for the three habitat groups were calculated from the multinomial regression model applied separately to each group. *P* values are for the hypothesis of no change in the basal fraction with *S*. The regression coefficient for *S* in the combined data, ignoring habitat, is -0.0219 ($P = 0.008$).

ed statistical analysis of their data, fearing that their webs were not a random sample from some defined population of webs. This is always a concern, but qualitative inferences from nonrandom samples are as risky as quantitative inferences, in our opinion.

The nature of the relationships of trophic fractions to *S* varies among habitat types (Figs. 1–3). In particular, phytotelmata and carrion-based webs seem to have patterns that are distinctly different from those in webs from the other habitat types. Statistical analysis cannot reveal whether these are fundamental ecological differences or methodological artifacts. For example, there is evidence that calculated trophic fractions are sensitive to the level of resolution of the nodes of the food web (Martinez 1993*b*). If workers in different habitat types use different degrees of trophic lumping and splitting, differences in the inferred

relationships of trophic fractions to *S* might well be a consequence.

Connectance also varies strongly with the number of species in the web (Fig. 4). A decrease in connectance with *S* has also been noted by other workers analyzing these data (Sugihara et al. 1989, Schoenly et al. 1991, Bengtsson 1994). The decrease in connectance with increasing *S* could be a consequence of biological constraints, or it could be a methodological artifact. For example, Paine (1988) suggested that connectance estimates are influenced by the extent of aggregation of species in nodes of the web and by the omission of links in large, complicated webs for “artistic convenience.”

Four habitat groups were statistically distinguishable with respect to their connectance patterns (Fig. 4). These differences, some of which were noted by

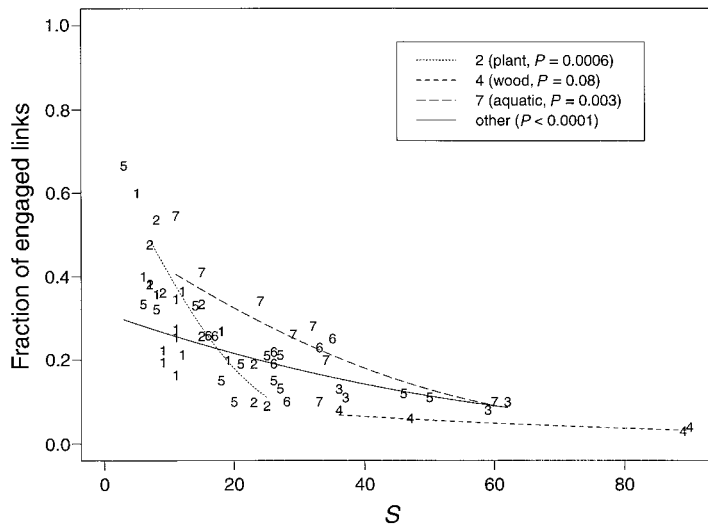


FIG. 4. Fraction of engaged links (number of links divided by $S[S - 1]/2$) vs. number of species *S* in the web. Plotted numbers indicate habitat type (see Table 1). Fitted lines for the four habitat groups judged to be distinct were calculated from the logistic regression model applied separately to each group. *P* values are for the hypothesis of no change in connectance with *S*. The regression coefficient for *S* in the combined data, ignoring habitat, is -0.030 ($P < 0.0001$).

Bengtsson (1994), could be due to morphological and behavioral differences among the species present in each habitat (Warren 1990), or they could be explained by variation in the methodology of ecologists working in different habitats (May 1983, Paine 1988). Whatever the explanation, it is clear that the patterns observed in large aggregates of food webs will be very much influenced by the mix of habitat types included in the data sets.

As evident from the above discussion, inferences based on collections of food webs—even the carefully assembled data of Schoenly et al. (1991)—are open to the challenges of systematic biases in the enumeration of species or trophic links associated with web size, organism type, and nature of the habitat. Yet, even in a set of 50 pelagic food webs assembled with consistent sampling and similar taxonomic resolution at all trophic levels (Havens 1992), there is strong evidence that trophic fractions vary with S (Martinez 1993a, Murtaugh 1994), suggesting that the scale dependence seen in the present data set may be ecologically meaningful. On the other hand, neither linear regression on transformed data (Havens 1992, Martinez 1993a) nor the logistic regression approach described above leads to the rejection of the constant connectance hypothesis for the pelagic webs. Thus, it is at least plausible that the negative trends in connectance with increasing S in the Schoenly et al. (1991) data are methodological artifacts.

Some claims of scale invariance of food web properties have been made in the absence of statistical analysis, and others may be based on inefficient use of the available data. By aggregating webs of different habitat or organism types, some studies may have overlooked important but conflicting patterns in web subsets, interpreting the variation as noise on top of an overall invariant pattern. For example, in the present data set, the overall variation of the intermediate fraction with S is nonsignificant ($P = 0.77$ from Table 2), but two of the three habitat subgroups show highly significant increases in the intermediate fraction with S (Fig. 2).

In each of the generalized linear models constructed here, the measure of residual variation known as the deviance was much larger than one would expect if the model were correct. That is, the probabilities of a species being in a particular trophic level or of a link being engaged vary substantially from web to web, even after adjustment for the number of species and the habitat type of the web. It could be that procedural idiosyncracies of the investigators studying different webs account for much of this variation; or there could be other important, unidentified factors influencing food web structure, such as the strength of the interactions among

species (see examples in Polis and Winemiller 1996). The existence of this unexplained variation in trophic fractions and connectance suggests that it may be difficult to identify laws governing food web properties over a wide variety of organisms and habitats.

ACKNOWLEDGMENTS

We thank D. Schafer for helpful statistical discussions, and J. Bengtsson, N. Martinez, and an anonymous reviewer for constructive comments on earlier versions of the manuscript.

LITERATURE CITED

- Becker, R. A., J. M. Chambers, A. R. Wilks. 1988. The new S language: a programming environment for data analysis and graphics. Wadsworth and Brooks/Cole, Pacific Grove, California, USA.
- Bengtsson, J. 1994. Confounding variables and independent observations in comparative analyses of food webs. *Ecology* **75**:1282–1288.
- Briand, F., and J. E. Cohen. 1984. Community food webs have scale invariant structure. *Nature* **307**:264–266.
- Cohen, J. E., F. Briand, and C. M. Newman. 1990. Community food webs: data and theory. Springer-Verlag, Berlin, Germany.
- Havens, K. 1992. Scale and structure in natural food webs. *Science* **257**:1107–1109.
- Hosmer, D. W., Jr., and S. Lemeshow. 1989. Applied logistic regression. John Wiley and Sons, New York, New York, USA.
- Martinez, N. D. 1992. Constant connectance in community food webs. *American Naturalist* **139**:1208–1218.
- . 1993a. Effect of scale on food web structure. *Science* **260**:242–243.
- . 1993b. Effects of resolution on food web structure. *Oikos* **66**:403–412.
- . 1994. Scale-dependent constraints on food web structure. *American Naturalist* **144**:935–953.
- May, R. M. 1983. The structure of food webs. *Nature* **301**:566–568.
- McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Chapman and Hall, London, UK.
- Murtaugh, P. A. 1994. Statistical analysis of food webs. *Biometrics* **50**:1199–1202.
- Paine, R. T. 1988. Food webs: road maps of interactions or grist for theoretical development? *Ecology* **69**:1648–1654.
- Polis, G. A. 1994. Food webs, trophic cascades and community structure. *Australian Journal of Ecology* **19**:121–136.
- Polis, G. A., and K. O. Winemiller. 1996. Food webs: integration of patterns and dynamics. Chapman and Hall, New York, New York, USA.
- Schoenly, K., R. A. Beaver, and T. A. Heumier. 1991. On the trophic relations of insects: a food web approach. *American Naturalist* **137**:597–638.
- Sugihara, G., K. Schoenly, and A. Trombla. 1989. Scale invariance in food web properties. *Science* **245**:48–52.
- Trexler, J. C., and J. Travis. 1993. Nontraditional regression analyses. *Ecology* **74**:1629–1637.
- Warren, P. H. 1990. Variation in food web structure: the determinants of connectance. *American Naturalist* **136**:689–700.
- Winemiller, K. O., and G. A. Polis. 1996. Food webs: what do they tell us about the world? Pages 1–22 in G. A. Polis and K. O. Winemiller, editors. Food webs: integration of patterns and dynamics. Chapman and Hall, New York, New York, USA.