

# Trophic links' length and slope in the Tuesday Lake food web with species' body mass and numerical abundance

DANIEL C. REUMAN\* and JOEL E. COHEN\*†

\*Laboratory of Populations, The Rockefeller University, Box 20, 1230 York Avenue, New York, NY 10021, USA; and

†Earth Institute at Columbia University, New York, NY 10027, USA

## Summary

1. Previously unnoticed patterns emerged in the lengths and slopes of trophic links of the food web of Tuesday Lake, Michigan, USA, when species were plotted on axes of log body mass (vertical) vs. log numerical abundance (horizontal). Link length was defined as equal to the number of orders of magnitude difference in mean body size between predator and prey plus the number of orders of magnitude difference in numerical abundance between predator and prey.

2. The average length over all trophic links was 6.2 [SD (standard deviation) = 2.8] orders of magnitude in 1984, and 5.8 (SD = 2.6) in 1986. Thus, for the average link, the ratio of the mean body mass of predator to prey, times the ratio of the numerical abundance of prey to predator, was about one million ( $10^6 \pm$ ). To a first approximation the typical predator was 10 times as long as its prey but 1000 times less numerically abundant. Link length distributions were approximately normal. Mean link lengths in 1984 and 1986 were not statistically different, but were more than an order of magnitude larger than the mean distance between all possible ordered pairs of species in each year. Ordered pairs of species that were between 9 and 11 orders of magnitude apart were the most likely to be links in both years.

3. The angle of a (non-cannibalistic) link was defined so that an angle of  $135^\circ$  indicates that the predator was larger than the prey by exactly as many orders of magnitude as the prey was numerically more abundant than the predator, and so that the biomass (mean body mass times numerical abundance) of predator and prey were equal. For all non-cannibalistic links, the median angle was  $132^\circ$  in 1984 and  $129^\circ$  in 1986. Both of these angles were significantly less than  $135^\circ$ . Angle deviated much from its median value only for short links. When the link was short or had the median angle, the predator species' biomass was probably greater than that of the prey species.

4. Several models of food web structure failed to reproduce the observed normal distribution of link lengths. Observed predation matrices with species ordered by body mass had links arranged in blocks suggesting functional groups. Models that incorporated this block structure reproduced the normal link length distribution, but a model that forced a normal link length distribution did not produce blocks. The cascade model explained the median angle of trophic links.

*Key-words:* allometry, body size, cascade model, food web, niche model, numerical abundance.

*Journal of Animal Ecology* (2004) **73**, 852–866

## Introduction

The combination of information on species' abundance and body mass with the traditional food web directed

graph is a powerful descriptive tool to characterize an ecological community (Ulanowicz 1984; Moore, deRuiter & Hunt 1993; deRuiter, Neutel & Moore 1995; Rott & Godfray 2000; Bersier, Banasek-Richter & Cattin 2002). Cohen, Jonsson & Carpenter (2003) and Jonsson, Cohen & Carpenter (2004) used average body mass ( $M$ ) and numerical abundance per unit of habitat ( $N$ ), attaching an ( $M$ ,  $N$ )-pair to each species in the food web (hereafter,

Correspondence: Joel E. Cohen, Laboratory of Populations, The Rockefeller University, Box 20, 1230 York Avenue, New York, NY 10021, USA. E-mail: cohen@rockefeller.edu

simply web) of Tuesday Lake, Michigan, USA. Unlike studies of single or a few predator–prey relations in simplified natural, experimental or theoretical webs, the studies by Jonsson, Carpenter and the present authors investigate the multivariate relations among body mass, numerical abundance and all predator–prey interactions in a well-defined natural community. The motivation is to provide a detailed, quantitative, community-wide context for the study of predator–prey relations, first in Tuesday Lake and then as a model for empirical studies in other natural communities.

The community-wide perspective on predator–prey relations yields many new empirical regularities, to be reported below. These empirical regularities are unfamiliar because the data from Tuesday Lake are, so far, unique in combining comprehensive data on a web, body mass and numerical abundance (see Materials and methods for details). These data were used to address familiar ecological questions concerning the importance of functional taxonomic groups.

One of the numerous new patterns reported here is that the length (defined below) of trophic links, when species are plotted on  $\log(M)$  vs.  $\log(N)$  coordinates, is approximately normally distributed. Moreover, the mean length is about equal to the community span (the number of orders of magnitude body-mass variation in the community plus the number of orders of magnitude numerical abundance variation in the community) divided by the mean number of links in food chains in the web. We hypothesize that this rough equality will hold in other webs as well.

The hope that predation matrices generated by the cascade model or niche model combined with the  $M$  and  $N$  distributions of Tuesday Lake would give normal link length distributions failed decisively. Analysis of these models and four new models indicates that species' average body mass and numerical abundance are not sufficient to account for the observed normal distribution of link lengths, without also taking account of trophic specialization associated with major taxonomic groups such as phytoplankton, zooplankton and fish. This specialization appears as rectangles in a body-mass-indexed predation matrix. The analysis indicates that taxonomy matters for feeding relations, beyond the very important effects of average body mass and numerical abundance. A new 'cascade model with functional groups' attempts to model webs with additional  $M$  and  $N$  data, and to combine functional specialization with other food web patterns to produce the observed normal distribution of link length.

Another new reported pattern is that the angles of trophic links from the positive  $x$ -axis when species are plotted on  $\log(M)$  vs.  $\log(N)$  coordinates have median slightly but significantly less than  $135^\circ$ . Because the species arrange themselves roughly in a linear pattern of slope about  $-1$  when plotted on these coordinates, it is to be expected that the median angle of links is roughly  $135^\circ$ . However, the median angle of links is also significantly less than the angle of the best-fitting line to species on

a  $\log(M)$  vs.  $\log(N)$  coordinates. The cascade model is sufficient to explain the observed median angles.

Normal link length distribution and many of the other statistical regularities found in Tuesday Lake, in addition to their direct biological significance, are also quantitative benchmarks to which future models of webs with  $M$  and  $N$  data should be compared.

The Tuesday Lake data cannot address questions of dynamics because the data describe only two points in time (1984 and 1986). These data also cannot identify the effects of the intervention in 1985 (removal of three fish species and introduction of another fish species) because no control lake is available.

## Materials and methods

### DEFINITIONS

All logarithms in this paper are base 10. Cohen *et al.* (2003) and Jonsson *et al.* (2004) suggested placing all species and trophic links in a web on axes with ordinate  $\log(M)$  and abscissa  $\log(N)$ . Then the  $l_1$  length of a link ( $r, c$ ) from prey (resource)  $r$  to predator (consumer)  $c$  is:

$$l_1 = |\log(M_c) - \log(M_r)| + |\log(N_c) - \log(N_r)| \\ = |\log(M_c/M_r)| + |\log(N_c/N_r)|, \quad \text{eqn 1}$$

the order of magnitude of the body size ratio between the predator and prey, plus the order of magnitude of the numerical abundance ratio between the predator and prey. The slope  $\alpha$  of a link ( $r, c$ ) is:

$$\alpha = \frac{\log(M_c) - \log(M_r)}{\log(N_c) - \log(N_r)}, \quad \text{eqn 2}$$

the order of magnitude of the body size ratio between predator and prey, over the order of magnitude of the numerical abundance ratio between predator and prey (which is typically negative). The angle  $\theta$  of a link ( $r, c$ ) is the angle in the counter-clockwise direction from a right-pointing horizontal line to the link considered as a vector from prey  $r$  to predator  $c$  in the  $(\log(N), \log(M))$  plane.

A trophic link that has an angle of  $135^\circ$  (equal to slope  $-1$ ) has predator and prey of equal biomass abundance. If the angle is less than  $135^\circ$  but greater than  $-45^\circ$ , then the biomass of the predator exceeds that of the prey. If the angle is greater than  $135^\circ$  but less than  $315^\circ$ , the opposite holds.

Taking population production and ingestion of a species to be proportional to  $NM^{0.75}$  (Peters 1983), one can show that a trophic link of slope  $-4/3$  has predator and prey species with equal production and ingestion. A slope that is less than  $-4/3$  with heavier predator indicates that the predator has greater ingestion and production than the prey, and a slope that is greater than  $-4/3$  with heavier predator indicates the opposite. Slope and angle are not defined for cannibalistic links where  $r = c$ .

Slope and angle are conceptually interchangeable, but there were statistical and mathematical circumstances

under which each was most appropriate, so both were used. For instance, angle interacts better with median and mean. Two links with angles  $89^\circ$  and  $91^\circ$  have mean angle  $90^\circ$ , but mean slope 0 because slope has a vertical asymptote at angle  $90^\circ$ . Slope is more useful than angle for certain linear regressions because the homoskedasticity assumptions of linear regression are more nearly satisfied for slope.

The  $l_2$  length of a link is the Euclidean length in the  $\log(M)$  vs.  $\log(N)$  plane, and  $l_1/l_2 = (|\alpha| + 1)/(\alpha^2 + 1)^{1/2}$ . For Tuesday Lake, where  $\alpha$  is usually close to  $-1$ ,  $l_1/l_2$  will be roughly  $2^{1/2}$ . A constant factor will not affect the trends examined, so  $l_1$  only is used because it, unlike  $l_2$ , has a clear biological interpretation. The intuitive understanding of the Euclidean distance applies to  $l_1$  because it differs from  $l_2$  only by a (very nearly) constant factor.

Each link in a web may be represented by a point in a 1-, 2-, 3-, or higher dimensional Euclidean space, depending on whether 1, 2, 3 or more quantitative attributes of the link are to be studied. The set of all links is represented by a cloud of such points. For example, below we study the one-dimensional distributions of each link's length and each link's angle (being careful of cannibalistic links).

Length, angle and other distributions are also well-defined for any ordered pair of species ( $a$ ,  $b$ ) in the  $(\log(N), \log(M))$  plane if one replaces  $c$  ('predator') with  $b$  and  $r$  ('prey') with  $a$  in eqn 1 and eqn 2 [again excepting the angle of 'cannibalistic' ordered pairs ( $a$ ,  $a$ )]. When discussing an ordered pair ( $a$ ,  $b$ ), species  $a$  will be called the prey and  $b$  the predator, even though there may have been no trophic relationship between  $a$  and  $b$ , or even if in reality  $a$  ate  $b$ . The set of links is contained in the set of ordered pairs, and there is no implied relationship between the body mass or numerical abundance of  $a$  and that of  $b$  in the ordered pair ( $a$ ,  $b$ ).

Species were divided into basal, intermediate and top (B, I and T) groups, allowing a division of ordered pairs into (B, B) (B, I) (B, T) (I, B) (I, I) (I, T) (T, B) (T, I) and (T, T) groups. This and several other divisions of species into groups permitted us to investigate whether these classifications of species were involved in statistical regularities in length and angle distributions. Any distribution of ordered pairs was thereby divided into subdistributions, one for each of these groups. Some (but not all) of these groups of ordered pairs contained links, and so subdistributions of links were also broken into groups. This procedure was carried out in other ways by starting with different initial groupings of species. The following groupings of species were used: the above BIT-grouping; a grouping that put species of similar average body mass  $M$  together (called the  $M$ -grouping); a grouping that put species of similar numerical abundance together (called the  $N$ -grouping); and the functional grouping into phytoplankton, zooplankton and fish species (called the PZF-grouping). In Tuesday Lake, the PZF-grouping was almost identical to the  $N$ -grouping.

All groupings are listed in Appendix S1 (see Supplementary material; also available on request from the authors), Table A1 and Table A2. The groups of the  $M$ -

grouping were called the H group, the S group and the L group, representing heavy, standard weight and light species. The groups of the  $N$ -grouping were called the R group, the U group and the C group, representing rare, uncommon and common species.

The trophic position of a species in a food chain is the number of species below it in the chain (so a species with no species below it has position 0). This definition is a slight modification of the definition of Jonsson *et al.* The trophic height of a species in a web is computed by collecting all chains that begin at the given species going strictly down through the web (i.e. from predator to prey at each step), but that do not contain more than one copy of the same species. One then takes the mean of the species' trophic position in all these chains. No chains including cannibalistic links are included in the mean, and no chains that go all the way around a cycle are included, although it is acceptable to go any part of the way around a cycle. The algorithm used by Jonsson *et al.* treated cycles in a different way, but produced results similar to this definition, other than a uniform difference of 1.

## DATA

The Tuesday Lake ( $M$ ,  $N$ )-enriched web of the non-littoral epilimnion contains a list of species and trophic relationships, average wet body mass (in kg) of individuals of each species, and numerical abundance of each species per  $\text{m}^3$  of water, measured in 1984 and again in 1986. In 1985, the three planktivorous fish species were removed and replaced by a single piscivorous fish species. Jonsson *et al.* (2004) described Tuesday Lake and presented the raw data and their limitations. Our analyses used unlumped webs and excluded isolated species (those not involved in any reported trophic relationship with any other species). There were six isolated species in both 1984 and 1986.

## METHODS

The  $M$ - and  $N$ -groupings were obtained by using a kernel density estimator of the distributions by species of  $\log(M)$  and  $\log(N)$ , respectively, similar to that described in Havlicek & Carpenter (2001) and Silverman (1986). The kernel density estimator is defined as the sum of normal probability density functions of common standard deviation  $w$ , centred at each species'  $\log(M)$  or  $\log(N)$  value. For large  $w$ , the resulting density functions had no local minima because the underlying normal density functions were so wide that they blurred together. As  $w$  decreased, local minima emerged. A kernel density estimator for  $w = 0.6$  shows the first two local minima (Appendix S1, Fig. A1), which were used to separate the groupings.

Computations were performed using the statistics toolbox of Matlab, version 6.5.0.180913a (R13). The Matlab statistics toolbox function 'ksdensity' was used for kernel density estimates. Linear regressions were done using 'regress'. Non-linear fitting used 'nlinfit'.

One-way ANOVAS used 'anova1'. Non-parametric testing was performed using 'ranksum' (Wilcoxon rank sum test). To determine whether the median of a non-normal distribution of  $n$  data points was different from a fixed value  $x$ , we let  $m$  be the minimum of the number of data points below  $x$  and the number above  $x$ . We then computed  $2\varphi_{n,0.5}(m)$  where  $\varphi_{n,0.5}$  is the cumulative distribution function (c.d.f.) of the binomial distribution with parameters  $n$  and  $1/2$ . This is the  $P$ -value of a two-tailed test with null-hypothesis that the median is equal to  $x$  and alternative hypothesis that it is not. We will refer to this procedure as the 'binomial test' (Siegel 1956). To determine whether a one-dimensional distribution is normal, the Jarque–Bera test (Jarque & Bera 1987) as implemented by 'jbtest', and the Lilliefors test (Lilliefors 1967) as implemented by 'lillietest', were used. The Lilliefors and Jarque–Bera tests are based on different characteristics of the data, so a distribution had to pass both tests to be considered normal. A distribution of data was considered to have passed a normality test with any  $P$ -value greater than 0.01 (rather than 0.05, because only a rough, qualitative description of the data was desired, and a  $P$ -value of more than 0.01 on both tests indicated that normality offered such a description).

The usage 'links (pairs)' occurring in a sentence will be used to denote that the sentence could be read using either the word links or the word pairs, producing a true statement in either case.

The collections of numbers that were tested for normality using the Jarque–Bera and Lilliefors tests were not independent because they were distributions of link (pair) lengths and angles, and a single species can be part of several links (pairs). This lack of independence means that one cannot interpret the  $P$ -values returned by these normality tests as probabilities. The Jarque–Bera  $P$ -values indicate how much the skew and kurtosis of the data differed from those of a normal distribution. The Lilliefors  $P$ -values indicate how much the sample c.d.f. differed from that of a normal distribution. These  $P$ -values should be regarded as descriptive statistics, rather than probabilities.

Cannibalistic links and pairs were necessarily excluded from analyses involving angle, as angle is not well defined for them. Angle distributions occur on a circle, not on a real line. One must break the circle to identify angle distributions with distributions on some part of the real line (necessary to apply the above statistical methods). For links, a large part of the circle always contained no values. Any point in this area could be chosen as the breaking point, and  $-135^\circ$  was chosen. For pairs, the circle was also broken at  $-135^\circ$ , where the fewest values occurred. Another approach was to consider only pairs for which the angle fell between  $45^\circ$  and  $225^\circ$ . Any pair  $(x, y)$  with  $x$  different from  $y$  has a corresponding pair  $(y, x)$  whose angle differs by  $180^\circ$ , so this approach eliminates a form of redundancy, but also excludes a few links.

Links or pairs whose slope was undefined, positive infinity or negative infinity were excluded from all analyses involving slope.

## MODELS

Models of webs usefully sharpen understanding of how observed patterns arise (Cohen *et al.* 1993). Link length distributions simulated by six models were investigated. Each model took as given the body-mass and numerical abundance distributions of Tuesday Lake, modelling only the selection of links from the set of ordered pairs. The models were the cascade model (Cohen 1990; Cohen, Briand & Newman 1990), an adaptation of the niche model of Williams & Martinez (2000), and four new models.

First, in the cascade model, the species index was interpreted as a rank ordering of body mass. The explicit limitation of the cascade model to trophic species was ignored here, although the cascade model was presented originally as a model of lumped webs (Cohen *et al.* 1990). Beginning with the body mass and numerical abundance distributions of unlumped species in Tuesday Lake from 1984, 269 links were randomly and uniformly chosen (the Tuesday Lake web had 269 links in 1984) for which the predator had a higher species index (body mass) than the prey. The process was repeated, choosing 264 links instead of 269 (5 links in Tuesday Lake were cannibalistic in 1984). The web for 1986 was simulated similarly. The biological content of this model is the assumption that any given species will eat, independently and with equal probability, any other species of lesser average body mass than itself. In theory, the cascade model can give rise to isolated species. However, 1000 simulated cascade predation matrices with either 50 species and 264 or 269 links, or 51 species and 236 or 241 links, never yielded more than four webs with isolated species (and never more than one isolated species per web).

Secondly, in the 'niche model' of Williams & Martinez (2000), the 'niche value' of each species was chosen uniformly on the interval (0,1). Their biological motivation was that a species cannot eat any species smaller than itself, but only those in a limited range of sizes; see details in Williams & Martinez (2000). The adapted version used here re-normalized the log body masses of the species to create their niche values. The re-normalization was linear, and sent the log of the smallest body mass to  $y$  and the log of the largest to  $x$ , where  $x$  was the maximum of  $n$  uniform random variables on (0,1) ( $n$  was the number of species in the model web) and  $y$  was the minimum of the same set of random variables. The original niche model eliminated isolated species and trophically identical species by deleting and replacing them. The adapted model allowed trophic duplicates because model results were to be compared with unlumped Tuesday Lake webs, but threw away simulated webs that contained isolated species because species could not be deleted individually (the niche values were determined in advance by the Tuesday Lake body mass distribution, which could not be changed).

A third, 'block equivalence' model ordered the species of Tuesday Lake by body mass, separately for

each year. The predation matrix was rearranged correspondingly, producing the same ( $M$ ,  $N$ )-web, annotated in a new way. Three blocks of links were apparent in the re-indexed predation matrices. They were outlined by hand from each of the predation matrices of 1984 and 1986. These blocks are described exactly in Appendix S1, Table A9 and Fig. A2. The model randomly permuted the entries in each block separately, to investigate whether block structure could explain the link length distribution of Tuesday Lake. The model does not explain what causes the blocks.

A fourth, 'diagonal equivalence' model started from Tuesday Lake ( $M$ ,  $N$ )-webs re-indexed by body mass as above. The model then permuted independently the entries of each diagonal of the resulting predation matrix to create a new ( $M$ ,  $N$ )-web, with the same body mass and numerical abundance distributions as the original. In a body-mass-indexed predation matrix, there is a rough correspondence between which diagonal an ordered pair is in and the length of that pair. On the other hand, the diagonal equivalence model disrupts the block structure of the body-mass-indexed predation matrices. The diagonal equivalence model seeks to determine whether preserving the distribution of links among the diagonals preserves the link length distribution of Tuesday Lake, in spite of disruption of the block structure.

A fifth 'cascade model with functional groups' added links with the following characteristics to a predation matrix, with equal probability. As before, species were ordered by body mass and the  $M$  and  $N$  distributions of Tuesday Lake were used. First, only links strictly above the main diagonal of the predation matrix were allowed (a 'size diet limit'). Secondly, a 'perceptual limit' was assumed to prevent a species from eating another species with body mass more than 6.8 orders of magnitude smaller. Thirdly, considering years independently, it was assumed that the first several species (by index) were primary producers (the first 19 in 1984 and the first 25 in 1986), and that the next several species (species 20–40 in 1984 and species 26–41 in 1986) only ate the first group, or were primary producers themselves. The numbers 19 and 25 are the numbers of species in 1984 and 1986 that were lighter than the respective lightest non-basal species. The numbers 40 and 41 are the indices of the heaviest species in 1984 and 1986 to eat only from the first group of species. The remaining species were allowed to eat anything within the confines of the size diet limit and the perceptual limit. This model postulates some biological causes of the block structure assumed in the block equivalence model.

A sixth 'forced link length distribution model' created upper-triangular, body-mass-indexed predation matrices. Link length distributions were forced to be similar to that of the links in the upper-triangular part of the body-mass-indexed Tuesday Lake predation matrix. The Matlab function 'ksdensity' was used to add together one normal probability density function of standard deviation 0.8 centred at the length of each pair in the

upper triangle of the body-mass-indexed predation matrix. Another density function was created in the same way for links in the upper-triangle, and the ratio of the latter density function to the former density function was computed. For each year, this quotient was used to produce model predation matrices by assigning a pair in the upper-triangle of a body-mass-indexed predation matrix to be a link with probability given by the quotient function upon plugging in the length of that pair. By construction, the link length distributions of ( $M$ ,  $N$ )-webs produced in this way were similar to those of Tuesday Lake. This model seeks to determine whether a block structure of the predation matrix is necessary for a model to generate link length distributions similar to those of Tuesday Lake.

The first five models were tested against link length data by using each to generate 5000 ( $M$ ,  $N$ )-webs, computing the Jarque–Bera and Lilliefors statistics of the link length distribution of each, and comparing the resulting distributions of 5000 statistics to the Jarque–Bera and Lilliefors statistics of the Tuesday Lake link length distributions. A few examples of the output of the sixth model were inspected visually for the presence of rectangular blocks in the simulated predation matrices. The Matlab code for all models is in Appendix S1.

The cascade model was tested against link angle data by using it to generate 5000 ( $M$ ,  $N$ )-webs, computing the median angle of each, and comparing the resulting distribution of median angles to the median angle of Tuesday Lake.

## Results

### THEORETICAL PREDICTIONS

#### Allometric relations

Ecological communities have frequently exhibited a rough allometric relationship:

$$\log(N) = \beta_1 \log(M) + \gamma_1. \quad \text{eqn 3}$$

Estimates of  $\beta_1$  have varied by community (Peters 1983; Peters & Wassenberg 1983; Peters & Raelson 1984; Blackburn & Gaston 1999; Leaper & Raffaelli 1999). The equation:

$$\log(M) = \beta_2 \log(N) + \gamma_2 \quad \text{eqn 4}$$

is not just an algebraic manipulation of eqn 3 if one obtains  $\beta_1$  and  $\gamma_1$  by minimizing the sum-of-squared-log( $N$ )-error, and  $\beta_2$  and  $\gamma_2$  by minimizing the sum-of-squared-log( $M$ )-error, because usually  $\beta_2$  differs from  $1/\beta_1$ . Substituting these equations into eqn 1 gives, respectively:

$$l_1 = (1 + |\beta_1|) |\log(M_c) - \log(M_r)| \quad \text{eqn 5}$$

$$l_1 = (1 + |\beta_2|) |\log(N_c) - \log(N_r)|. \quad \text{eqn 6}$$

These results predict what to expect for the three-dimensional distributions ( $\log(M_r)$ ,  $\log(M_c)$ ,  $l_i$ ) and ( $\log(N_r)$ ,  $\log(N_c)$ ,  $l_i$ ) both over ordered pairs and over links.

Suppose that a species grouping contains a group G in which  $\log(M)$  varies little. Substituting the mean value of  $\log(M)$  over the group G for  $\log(M_c)$  in eqn 5 gives a prediction of  $l_i$  for links (pairs) with predator in G. The accuracy of the prediction depends on how accurately the mean of  $\log(M)$  approximates the actual values of  $\log(M)$  for all members of G. The accuracy of the prediction also depends on the extent to which the shape (as opposed to the centre) of the distribution of the values of  $\log(M_c)$  across all links (pairs) with predator in G and prey with fixed mass  $M_r$  is independent of  $\log(M_r)$ . The most accurate predictions will occur when the distribution of values of  $\log(M_c)$  across all links (pairs) with predator in G and prey with fixed mass  $M_{r1}$  is just a translation of the analogous distribution using fixed prey mass  $M_{r2}$ . Heteroskedastic data plots will arise if departures from this assumption occur (see Results). Such departures can cause the theoretical predictions just developed to be inaccurate.

Substituting eqn 4 into eqn 2 gives:

$$\alpha = \beta_2$$

eqn 7

for the slope. This result predicts the three-dimensional distributions ( $\log(M_r)$ ,  $\log(M_c)$ ,  $\alpha$ ) and ( $\log(N_r)$ ,  $\log(N_c)$ ,  $\alpha$ ) over links (pairs).

### Clumping

Jonsson *et al.* (2004, their Fig. 4) plotted all links in Tuesday Lake on  $\log(M_r)$  vs.  $\log(M_c)$  axes. The points on their graph separated roughly into clumps. Figure 1 shows links plotted on the same axes with ordinate and abscissa exchanged, but using different markers according to the  $N$ -grouping. Clumping of groups is evident, and was also evident when the other groupings were used. The clumps were delineated more precisely under some groupings than under others. These clumps affected many other distributions.

Clumps tended to be rectangular. After drawing the line  $y = x$  on the axes, it was always possible to draw another line of slope 1 so that every clump sat roughly between

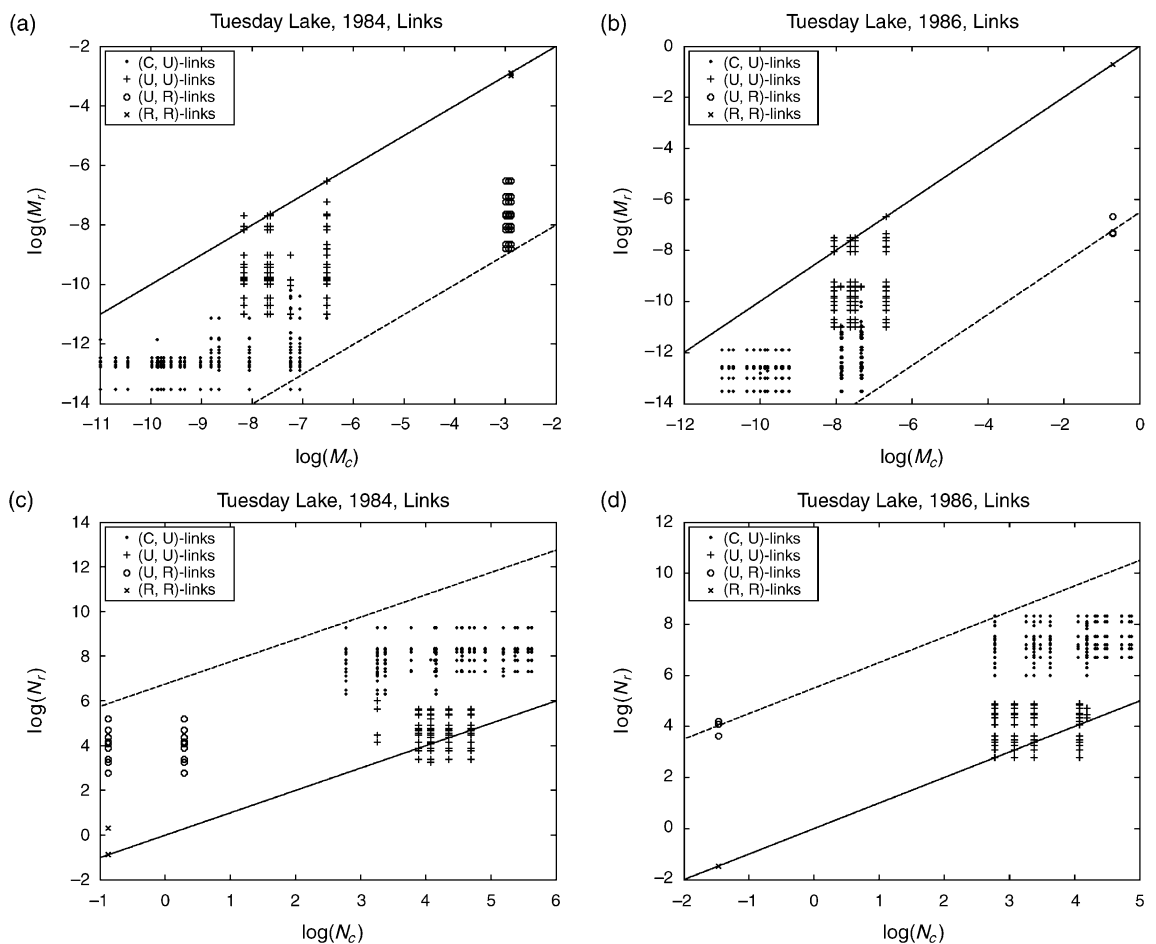


Fig. 1. Trophic links of Tuesday Lake in 1984 (a, c) and 1986 (b, d) by average body masses (a, b) and numerical abundances (c, d) of predator and prey species. Trophic links clumped into approximate rectangles associated with numerical abundance. The numerical abundance grouping was the same as the functional grouping into phytoplankton, zooplankton and fish in 1986, and different by one species in 1984. The numerical abundance groups are R, U and C for rare, uncommon and common. Links are grouped using (A, B), where A is the group of the prey and B is the group of the predator. Most links fell between the (solid) diagonal line  $y = x$  and a second (dashed) line with slope 1 which delimits the extent of links.

these two lines, with one of its corners on one of the lines. The second line of slope 1 was roughly invariant with respect to year. Predators usually did not prey on other species too many orders of magnitude smaller or more abundant. Large gaps in a single clump (as are apparent, for instance, in the clump consisting of '+' signs in 1986 in Fig. 1b) were due typically to gaps in the plot of ordered pairs on  $\log(M_p)$  vs.  $\log(M_p)$  axes, which in turn came from gaps in the distribution of  $\log(M)$  of species. Rectangular clumps of links plotted on  $\log(N_p)$  vs.  $\log(N_p)$  axes also occurred, as one would expect as the negative correlation between  $\log(M)$  and  $\log(N)$  was strong. The rectangles on these axes also fell roughly between the line  $y = x$  and another line of slope 1, with at least one corner of each rectangle on one of these lines. Because species in Tuesday Lake fell near a line of slope  $\beta_2$  (which was close to -1) in the  $\log(M)$  vs.  $\log(N)$  plane, one would expect long links to deviate less in slope from  $\beta_2$  than short links. One may also expect the median angle of all links to approximate  $\beta_2$ . Under any of the species groupings mentioned previously, species in the same group tended to clump together along this line of slope  $\beta_2$ . Different groups tended to be located at different places along the line. Therefore, one would expect only links in which the predator and prey are from the same grouping to deviate significantly from slope  $\beta_2$  as these were, to a large degree, the only short links. One would expect such a phenomenon when using any grouping that is fairly well correlated to  $\log(M)$  or  $\log(N)$ .

COMPARISON OF THEORY TO DATA

Descriptive statistics

Table 1 summarizes the basic descriptive statistics of the length, slope and angle of links and ordered pairs in 1984 and 1986. Link lengths were normally distributed, so parametric tests were used, and the mean was used as the measure of centre. Pair length distributions and angle and slope distributions were not normal, so non-parametric tests and the median were used. Dis-

tributions of pair angles are considered only in the range of angles between 45° and 225° (see Methods). For all links, the average  $\pm$  standard deviation of length was  $6.2 \pm 2.8$  orders of magnitude in 1984 and  $5.8 \pm 2.6$  in 1986. Thus, for the average link from a prey species to a predator species, the ratio of the mean body mass of predator to prey  $\times$  the ratio of the numerical abundance of prey to predator was about one million ( $10^6$ ), assuming that the predator had larger average body mass and lower numerical abundance than its prey. The mean length of links was more than two orders of magnitude larger than the median length of all ordered pairs of species, in 1984 and in 1986. The link length difference between years was not significant at the 5% level according to a one-way ANOVA. The pair length distributions in each year could be distinguished at the 1% level using the Wilcoxon rank sum test. The median angles of non-cannibalistic links were 132° and 129° in 1984 and 1986, respectively. The binomial test distinguished these median angles from 135° at the 1% level in both years. The binomial test also distinguished these median angles from the angles corresponding to the slopes  $\beta_2$  of the species linear regression lines (Table 2). The link angle distributions in 1984 and 1986 were not distinguishable at the 5% level using the Wilcoxon rank sum test. Links and ordered pairs were shorter in 1986 than in 1984, and links had smaller median angles. These differences were significant for pairs, but not for links. Thus, on average, in both links and pairs, predators and prey were closer in 1986 than in 1984 in the plane of log body mass and log numerical abundance. The ratio of predator biomass to prey biomass in links was typically larger in 1986 than in 1984.

Regression coefficients of allometric relations

Linear regression of the Tuesday Lake data, including only non-isolated species, gave values (Table 2) for the coefficients  $\beta_1$  in eqn 3 and  $\beta_2$  in eqn 4. Jonsson *et al.* (2004) and Cohen *et al.* (2003) included all species, not

**Table 1.** Basic descriptive statistics of links and ordered pairs in the Tuesday Lake food web. SD = standard deviation. Length, slope and angle are defined in the text. In entries of the form  $x/y$ ,  $x$  is the relevant quantity for all links or pairs and  $y$  is the relevant quantity for only non-cannibalistic links or pairs. Cannibalistic links were excluded from slope and angle calculations, and infinite values were also excluded from slope calculations. Angle calculations with ordered pairs included only pairs with angles between 45° and 225°

	Links			Ordered pairs		
	mean	SD	median	mean	SD	median
1984	(269 observed, 5 cannibalistic)			(2500 observed, 50 cannibalistic)		
Length	6.1777/6.2947	2.8246/2.7186	6.5049/6.5403	5.0467/5.1497	4.2257/4.2060	3.9220/4.0417
Slope	-1.1585	6.7734	-1.0278	-0.9689	6.0259	-0.9825
Angle	125.1476	26.2328	131.5118	133.4167	30.2410	132.5952
1986	(241 observed, 5 cannibalistic)			(2601 observed, 51 cannibalistic)		
Length	5.7724/5.8947	2.5717/2.4557	6.0313/6.1294	4.3756/4.4631	3.6593/3.6425	3.4210/3.5019
Slope	-0.8625	10.3901	-1.1300	-0.6919	5.3494	-0.8753
Angle	121.6181	32.4822	128.9777	133.9004	34.8867	132.6054

**Table 2.** Slope coefficients  $\beta_1$  and  $\beta_2$  of linear regressions between log numerical abundance  $\log(N)$  and log average body mass  $\log(M)$  of non-isolated species in Tuesday Lake in  $\log(N) = \beta_1 \log(M) + \gamma_1$  and  $\log(M) = \beta_2 \log(N) + \gamma_2$ . The values of the squared correlation coefficient  $R^2$  and the probability  $P$  that the slope differs from 0 by chance alone are the same for both regressions. The 99% confidence intervals for the slopes are in parentheses

	$\beta_1$	$\beta_2$	$R^2$	$P$
1984	-0.8413 (-0.9764, -0.7062)	-1.0141 (-1.1770, -0.8513)	0.8532	< 0.001
1986	-0.7461 (-0.9065, -0.5857)	-1.0191 (-1.2381, -0.8000)	0.7603	< 0.001

**Table 3.** The means and standard deviations of  $\log(M)$  and  $\log(N)$  and the number of species in each group from the  $M$ - and  $N$ -groupings for 1984 and 1986. The other groupings are in Appendix S1, Table A3

Group	No. species	log( <i>M</i> )		log( <i>N</i> )	
		mean	SD	mean	SD
1984					
<i>M</i> -grouping					
H (heavy)	3	−2.9390	0.0535	−0.4870	0.6768
S (standard)	29	−9.4276	1.3074	5.0171	1.3231
L (light)	18	−12.2999	0.5013	7.6413	0.6904
<i>N</i> -grouping					
R (rare)	3	−2.9390	0.0535	−0.4870	0.6768
U (uncommon)	23	−9.1147	1.2861	4.4772	0.8363
C (common)	24	−11.8817	0.8712	7.5027	0.6823
1986					
<i>M</i> -grouping					
H (heavy)	1	−0.7093	0.0000	−1.4685	0.0000
S (standard)	9	−7.5589	0.4161	3.4350	0.5070
L (light)	41	−11.3105	1.1361	6.0884	1.4355
<i>N</i> -grouping					
R (rare)	1	−0.7093	0.0000	−1.4685	0.0000
U (uncommon)	21	−8.9697	1.3518	3.8412	0.6568
C (common)	29	−11.8412	0.8431	6.8922	0.7181

only non-isolated species, and obtained similar but not identical values of  $\beta_1$  and  $\beta_2$ .

#### Constancy within groups

The means and standard deviations of  $\log(M)$  and  $\log(N)$  and the number of species in each group are shown in Table 3 for the  $M$ - and  $N$ -groupings and for other groupings in Appendix S1, Table A3.

To explain linear subrends such as those found in Fig. 2, we substituted into eqn 5 to obtain eqn 8 and eqn 9. To predict the  $l_1$  length of links and ordered pairs as a function of  $\log(M_r)$  when the predator is in the H group, one approximates the value of  $\log(M)$  for the H group by the mean value, obtaining:

$$l_1 = 1.7461 | -0.7093 - \log(M_r) | \quad \text{eqn 8}$$

in 1986 from eqn 5. Because there was only one species in the H group in 1986, the approximation by the mean of the  $\log(M)$  is exact in this case. Approximating the S group by its mean gives:

$$l_1 = 1.7461 | -7.5589 - \log(M_r) | \quad \text{eqn 9}$$

in 1986 as a predictor of how the length of links (pairs) for which the predator is in the S group should vary as

a function of  $\log(M_r)$ . The approximations underlying this equation were not as good as those for the H group. Similar results for 1984 are shown in Table 4.

One can also substitute for  $\log(N_c)$  in eqn 6 if  $\log(N)$  varied little in the group G. The approximations involved were better for the R group than for the U group (Table 4).

#### Three-dimensional distributions

Link (pair) length varied as a function of the body mass of both consumer and resource, or as a function of the numerical abundance of both consumer and resource, as predicted by theory in both 1984 and 1986. Specifically, when the non-linear equations

$$l_1 = | a \log(M_c) + b \log(M_r) | \quad \text{eqn 10}$$

$$l_1 = | a \log(N_c) + b \log(N_r) | \quad \text{eqn 11}$$

were fitted to the data ( $\log(M_r)$ ,  $\log(M_c)$ ,  $l_1$ ) and ( $\log(N_r)$ ,  $\log(N_c)$ ,  $l_1$ ) for links (pairs), using the Matlab non-linear fitting routine 'nlinfit', the fitted equations explained 75–93% of the variance of the data from the mean, and the theory eqn 5 and eqn 6 explained almost as much (Appendix S1, Table A4). The variables  $\log(M_c)$  and  $\log(M_r)$  were a better predictor of  $l_1$  than were  $\log(N_c)$  and  $\log(N_r)$ .

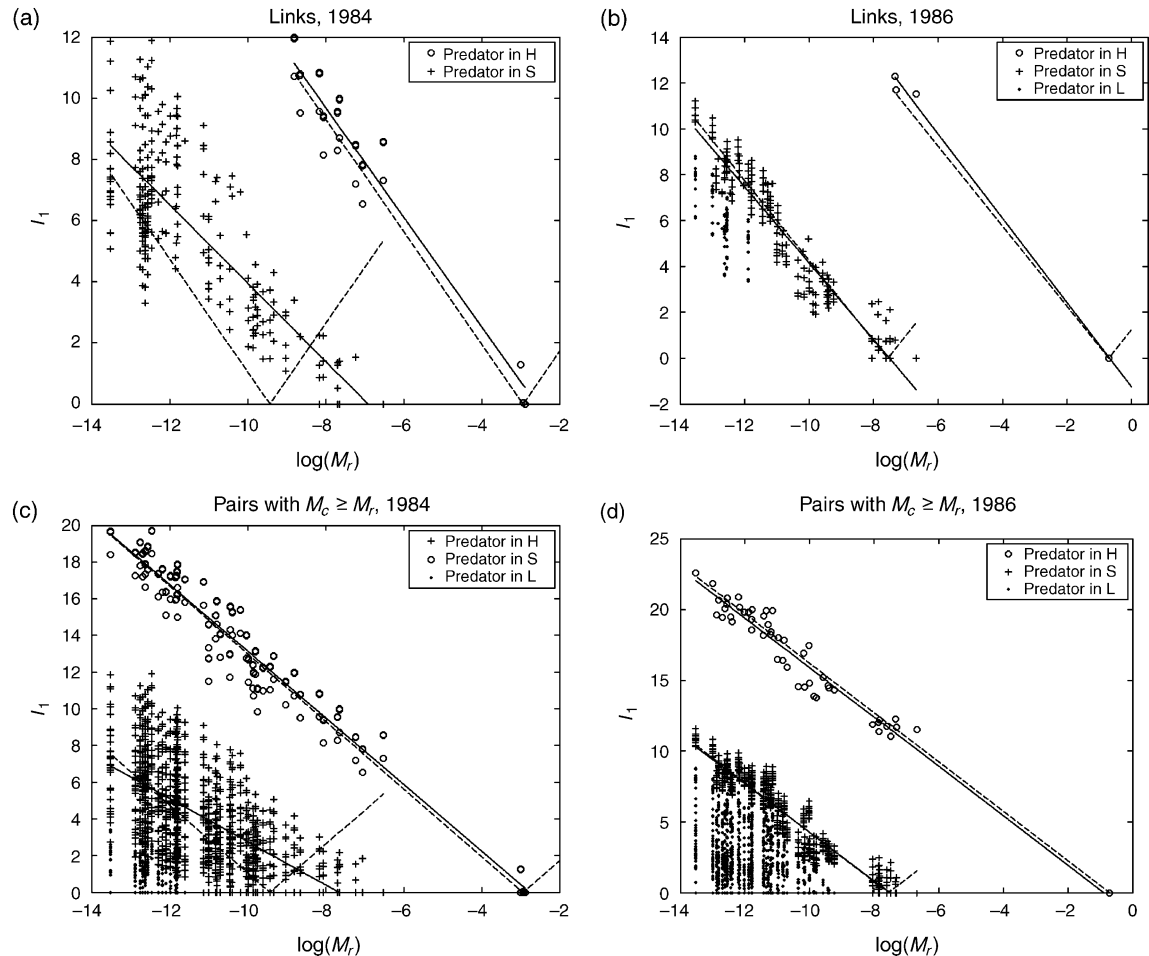


Fig. 2. The l1 length as functions of resource's log average body mass in Tuesday Lake, for trophic links (a, b) and ordered pairs of species with heavier consumer than resource (c, d), in 1984 (a, c) and 1986 (b, d). Markers show a grouping of predator or consumer species into categories of similar average body mass. H, S and L groups represent heavy, standard weight, and light. Solid lines are best fits to O and + symbols, respectively. Dashed lines are the corresponding predictions of the Theoretical Predictions section.

Table 4. Results of substituting for log(Mr) in l1 = (1 + |β1|) | log(Mr) - log(Mr) | or for log(Nr) in l1 = (1 + |β2|) | log(Nr) - log(Nr) | if log(M) (or log(N), respectively) varies little or is constant in the group G of predators, in 1984 and 1986 for G equal to the H and S groups of the M-grouping or the R and U groups of the N-grouping

Year	H (heavy) group	S (standard) group
1984	$l_1 = 1.8413   - 2.9390 - \log(M_r)  $	$l_1 = 1.8413   - 9.4276 - \log(M_r)  $
1986	$l_1 = 1.7461   - 0.7093 - \log(M_r)  $	$l_1 = 1.7461   - 7.5589 - \log(M_r)  $
	R (rare) group	U (uncommon) group
1984	$l_1 = 2.0141   - 0.4870 - \log(N_r)  $	$l_1 = 2.0141   4.4772 - \log(N_r)  $
1986	$l_1 = 2.0191   - 1.4685 - \log(N_r)  $	$l_1 = 2.0191   3.8412 - \log(N_r)  $

Length varied with the trophic heights of predator and prey, as would be expected for ordered pairs with heavier predator than prey, but for links, unexpectedly, did not increase significantly with the trophic height of the consumer. To model the distribution (Hr, Hc, l1) (where Hr and Hc represent the trophic height of the prey and predator, respectively), restricted to ordered pairs for which Mc > Mr, and restricted to links, the equation:

l1 = aHc + bHr + c eqn 12

was fitted to the data (Table 5). Over ordered pairs with Mc > Mr, l1 increased with increasing Hc, and decreased

with increasing Hr, as expected, because greater trophic height was associated with higher body mass, and therefore corresponded to species higher along the line represented by eqn 4. Moving a predator up, and prey down, along this line increased link length. For links, the relationships were far less clear, particularly the relationship between l1 and Hc. The 99% confidence intervals for the coefficient of Hc contained 0 in both 1984 and 1986. Other measures of trophic height all provided relationships less clear than the above. When the slope of a link α was regressed linearly on log(Mr) and log(Mc), the 99% confidence intervals for the slope coefficients of log(Mr) and log(Mc) contained

**Table 5.** Slope coefficients resulting from fitting  $l_1 = aH_c + bH_r + c$  to the data. The heading 'Pairs, heavy pred.' refers to pairs where predator is strictly heavier than prey. The 99% confidence intervals for the fitted coefficients are in parentheses.  $H$  is trophic height

	$a$	$b$	$R^2$	$P$
Links				
1984	0.520 (−0.051, 1.090)	−1.722 (−2.479, −0.964)	16.6%	< 0.001
1986	−0.422 (−0.892, 0.047)	−1.456 (−2.050, −0.863)	43.1%	< 0.001
Pairs, heavy pred.				
1984	2.643 (2.518, 2.768)	−2.951 (−3.228, −2.674)	70.6%	< 0.001
1986	2.301 (2.174, 2.429)	−2.622 (−2.910, −2.334)	63.0%	< 0.001

0 in both years, over both non-cannibalistic links and ordered pairs with heavier predator. The median slope of links in 1984 was −1.03 and in 1986 was −1.13, and the median slopes of pairs were −0.983 and −0.875, respectively. These results were reasonably close to the predictions of theory (−1.01 and −1.02, by eqn 7 and Table 2), and the theoretical values fell within 99% confidence intervals of the  $z$ -intercept of the linear regressions of slope vs.  $\log(M_r)$  and  $\log(M_c)$  fitted to link (pair) data. Regression results are in Appendix S1, Table A5.

Similar linear regressions of  $\alpha$  vs.  $\log(N_r)$  and  $\log(N_c)$  for links and for pairs with heavier predator (Appendix S1, Table A5) showed that zero fell in all 99% confidence intervals of slope coefficients except two (both slope coefficients of pairs with heavier predator in 1986), and these intervals almost contained zero. The theoretically predicted values of  $\alpha$  always fell within 99% confidence intervals of the  $z$ -intercepts of the fitted planes.

Similar linear regressions with angle in place of slope are also possible. However, angles plotted against  $\log(M_c)$  and  $\log(M_r)$ , or against  $\log(N_c)$  and  $\log(N_r)$ , are not nearly homoskedastic. The homoskedasticity assumption is more nearly, although not perfectly, satisfied for slope. Visual examination of scatter plots of slope confirms that the linear regressions above describe the slope data well.

#### Two-dimensional distributions

The larger the body mass of the prey, the shorter the length of the link (pair) for any given group of predator body mass, and the larger the numerical abundance of the prey, the longer the length of the link (pair) for any given group of predator numerical abundance (Fig. 2 for  $(\log(M_r), l_1)$ , Appendix S1, Fig. A3 for  $(\log(N_r), l_1)$ , and Appendix S1, Table A5 for the coefficients for the theory and the best-fit, and 99% confidence intervals for the best-fit). For links (pairs) for which the predator was in the H group or the R group, theoretically predicted slope coefficients were always within the 99% confidence intervals of the fitted values. Theoretically predicted  $y$ -intercept values were also within 99% confidence intervals of the fitted results for the H and R groups, except for pairs when  $l_1$  was regressed on  $\log(N_r)$ .

Theoretical predictions sometimes corresponded less well to fitted values for links (pairs) with predator in the S group for two reasons. First, the approximation of  $\log(M_c)$  by the mean of  $\log(M_c)$  across the S group was

rough in some cases (for instance, links or pairs in 1984). Secondly, the shape of the distribution of  $\log(M_c)$  across links (pairs) with given  $\log(M_r)$  was not always independent of  $\log(M_r)$ . This lack of independence led to the heteroskedastic distributions of '+' signs found in Fig. 2a,c. The predictions of theory were good in 1986 for links and pairs, but not in 1984.

For the same reasons, all theoretical predictions involving links (pairs) with predator in the U group failed to correspond closely with fit.

Long links were much more likely than short links to have predator and prey with equal (or nearly equal) biomass abundance. As theory predicted, when  $l_1$  was plotted as a function of  $\theta$ , the angle of a link (pair) (e.g. Fig. 3 for links, and for pairs with angle between 45° and 225°, using the  $N$ -grouping), only short links (pairs) deviated much from the median angle, and the only links (pairs) that deviated much from the median angle were links (pairs) in which the predator and the prey were in the same group. The only short links were (R, R) links and (U, U) links. The Wilcoxon rank sum test was applied to the distribution of angles in which predator and prey  $N$ -groups were the same, vs. angles with predator and prey from different  $N$ -groups. The distributions could be distinguished with  $P < 0.001$  in 1984 and 1986, both for links and for pairs for which the angle was between 45° and 225°. Figure 3a,b also shows a lack of symmetry of the link ( $\theta, l_1$ ) distributions about the median angle (which is the angle at which long links tended to occur). In both 1984 and 1986, nearly all shorter links had angles less than about 135°, although a few had angles greater than 135°. This asymmetry is not surprising, because links of a fixed length that have angle less than 135° have more predator–prey size difference than links of the same length that have slope greater than 135°. By contrast, for pairs, the distribution of angles was symmetric.

#### One-dimensional distributions

One-dimensional distributions of angles (Table A7) and lengths (Table A8) were tested for normality.

Many patterns of normality were discovered in angle distributions of ordered pairs. These are reported in Appendix S1. The distribution of angles over non-cannibalistic links was not normal in either year. Its left tail was too long. The only coherent patterns that emerged when considering links within certain groupings is that (Z, Z) (S, S) (U, U) and (I, I) links that had

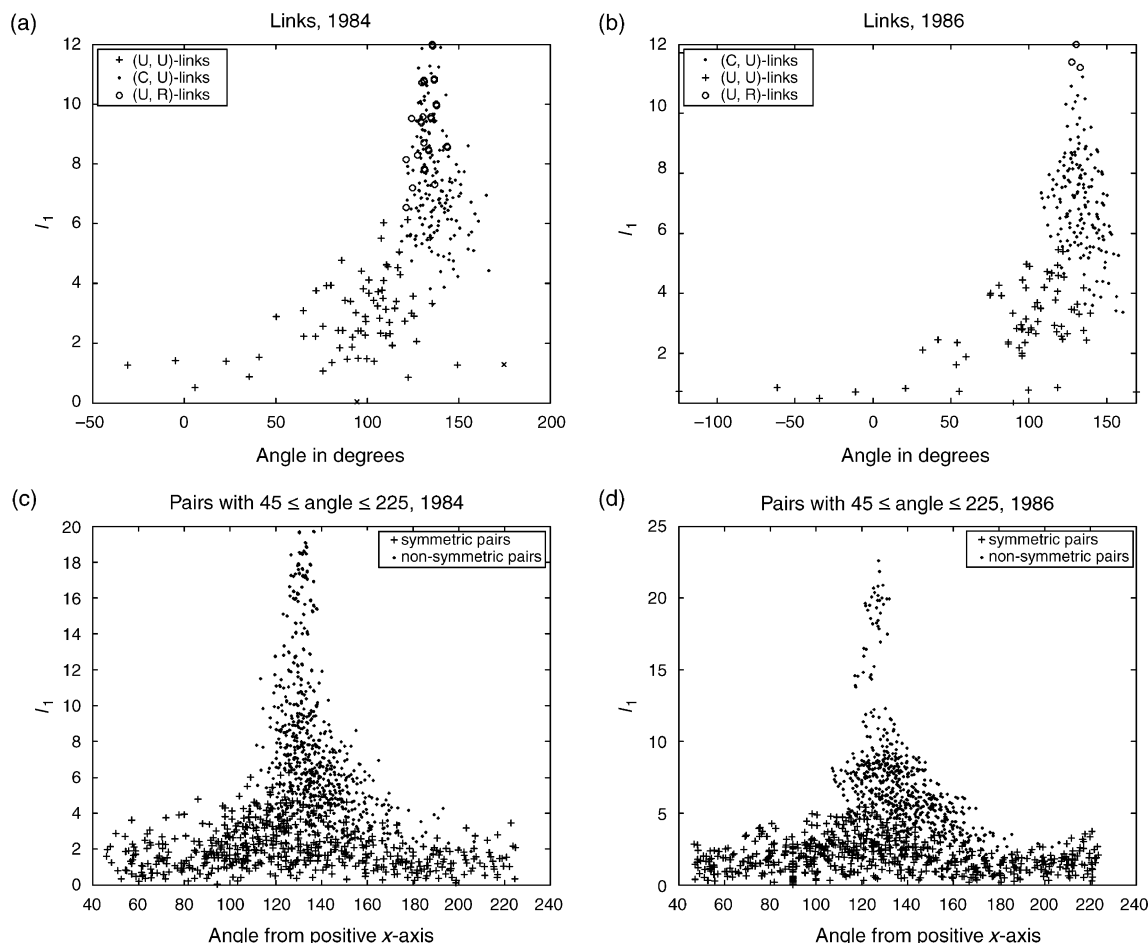


Fig. 3. The  $l_1$  length of each link or pair as a function of the angle between the link or pair and the positive  $x$ -axis, for trophic links (a, b) and ordered pairs of species (c, d) of Tuesday Lake in 1984 (a, c) and 1986 (b, d). Pairs in (c) and (d) are limited to those with angle from the positive  $x$ -axis lying between  $45^\circ$  and  $225^\circ$ . Markers show grouping of species into categories of similar numerical abundance. R, U and C groups represent rare, uncommon and common. Only short links or pairs deviate much from the mean angle. The short links or pairs are all from symmetric groups.

angles between  $45^\circ$  and  $225^\circ$  had normally distributed angles independent of year.

The one-dimensional distribution of  $l_1$  length was not normal over all ordered pairs. It leaned heavily toward zero, perhaps resembling a normal distribution truncated at zero, in both years (see Fig. 4 for 1984). This possibility was not tested statistically because we are not aware of a composite test for truncated-at-zero normality. Other distributions arising below from  $l_1$  also seemed, qualitatively, to be truncated normal. It may be worth designing a test of truncated-at-zero normality if other webs have qualitative patterns similar to those of Tuesday Lake.

The same results were true of the distribution of lengths over the set of pairs containing all cannibalistic pairs, and one of each of  $(x, y)$  and  $(y, x)$  for all  $x$  different from  $y$ .

Link lengths were normally distributed in both 1984 and 1986 (Fig. 4 for 1984). Against the backdrop of a highly non-normal distribution of pairs that was biased heavily toward short pairs, this normality emphasizes that short links are not preferred.

The  $l_1$  distributions of links over any group arising from the  $N$ -grouping were normal in 1984 (although

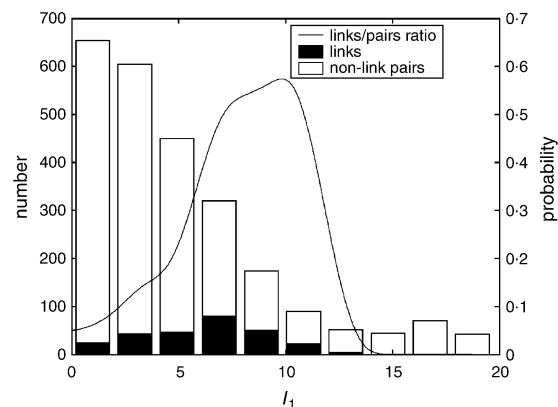


Fig. 4. Frequency distributions of lengths of all pairs and of all links in 1984 (histogram – left axis), and the quotient of the kernel density estimates of these distributions with window width 0.8 on the  $l_1$  scale (smooth curve – right axis). The frequency distribution of pairs is highly non-normal. The frequency distribution of lengths of links is approximately normal. The pairs most likely to be links have lengths between about 9 and 11 orders of magnitude. Similar phenomena occurred in 1986.

**Table 6.** Comparison of the Jarque–Bera and Lilliefors statistics of the link length distribution of Tuesday Lake in 1984 and 1986 with those of ( $M$ ,  $N$ )-webs built from the body mass and numerical abundance distributions (in 1984 and 1986, respectively) of Tuesday Lake using the cascade model. The cascade model does not produce cannibalistic links. Tuesday Lake link length distribution statistics were computed both including and excluding cannibalistic links, in the ‘cann.’ and ‘no cann.’ columns, respectively. In 1984 there were 269 links, five cannibalistic. In 1986 there were 241 links, five cannibalistic. The cascade model with all combinations of connectance parameters (264 and 269 in 1984 and 236 and 241 in 1986) was used to produce 5000 webs for each choice of parameters. The minimum, maximum and mean Jarque–Bera and Lilliefors statistics of the resulting link length distributions are shown. In every case, all 5000 model statistics were larger than the corresponding Tuesday Lake statistic

	Model simulations							
	Data		264 (1984) or 236 (1986) links			269 (1984) or 241 (1986) links		
	cann.	no cann.	min.	mean	max.	min.	mean	max.
Jarque–Bera								
1984	6.943	6.622	49.013	116.285	282.238	50.074	119.212	341.449
1986	3.452	2.520	25.628	382.966	1061.458	22.868	391.750	1040.741
Lilliefors								
1984	0.0524	0.0472	0.0930	0.1301	0.1818	0.0936	0.1299	0.1797
1986	0.0515	0.0462	0.0839	0.1266	0.1907	0.0816	0.1263	0.1740

there were too few data in (R, R) to draw conclusions) and in 1986 (although (R, R) and (U, R) had too few data to draw conclusions). Ordered pair distributions were normal in both years for non-symmetric groups (those of the form (A, B) where A and B are different groups of species), except for the (C, U) and (U, C) groups. The species-level groups C and U were close together in the  $\log(N)$  vs.  $\log(M)$  plane, and therefore may have behaved in some respects as one large group. No other pair of two different species-level groups was so close. The symmetric groups (C, C) and (U, U) could have had truncated normal distributions in both years. There were too few data in the (R, R) group to draw conclusions. Overall these data suggest that distributions of link and pair lengths were always truncated-at-zero-normal, both over all links (pairs), and within groups of the N-grouping. Normal distributions with mean many standard deviations from 0 appear to be normal when truncated at 0, as in any link distribution, and any pair distribution from a nonsymmetric group. For distributions over pairs from symmetric groups, truncation had a more visible effect. Groups of the form (A, B) where A and B are different but close on the  $\log(N)$  vs.  $\log(M)$  plane may have deviated slightly from this pattern.

#### *Modelling results on normality of link length distributions*

To illuminate the mechanisms that produced the observed normality of the link length distribution in the Tuesday Lake web, the normality of the link lengths of webs produced by the first five models described in Materials and methods was investigated. The cascade model did not produce link length distributions that resembled the Tuesday Lake link length distributions (Table 6). These results are extremely statistically significant, but are also expected. Cascade model link length distributions were expected to look like ordered pair length distributions because the cascade model

chooses links uniformly from the set of pairs with heavier predator. A variant of the cascade model for which each pair with heavier predator was assigned link status with equal probability would produce very similar results.

The adapted niche model also failed to produce link length distributions with Jarque–Bera and Lilliefors statistics similar to those of the Tuesday Lake webs (Table 7). These results are also extremely statistically significant. The diagonal equivalence model failed in the same way (Table 7), with a similar level of significance. The block equivalence model, however, produced statistics similar to those of the Tuesday Lake data, as did the cascade model with functional groups (Table 7).

The quotient of density functions computed for the forced link length distribution model is shown for 1984 in Fig. 4. The 1986 quotient was very similar. In both years, the pairs most likely to be links had lengths between about 9 and 11 orders of magnitude. By construction, the link length distributions of ( $M$ ,  $N$ )-webs produced by this model were similar to those of Tuesday Lake. However, predation matrices produced in this way did not exhibit the blocks found in Appendix S1, Fig. A2, although they did exhibit a rough perceptual limit.

Comparing the results from the first four models suggested that the block structure of the predation matrix in the block equivalence model was important for the resulting normal distribution of link lengths. The ‘cascade model with functional groups’ suggested that a size diet limit, a perceptual limit, and the imposition of functional groups were sufficient to approximate the blocks of trophic links that produced a normal distribution of link length. The ‘forced link length distribution model’ showed that blocks represent additional structure beyond that inherent in a normal link length distribution, since blocks led to normality, but not vice versa.

The cascade model produced webs with median angle comparable to those of Tuesday Lake. In 1984, 6.7% of the

**Table 7.** Comparison of the Jarque–Bera and Lilliefors statistics of the link length distribution of Tuesday Lake in 1984 and 1986 with those of ( $M, N$ )-webs built in four ways (see Methods): A: from the body mass and numerical abundance distributions (in 1984 and 1986, respectively) of Tuesday Lake by using an adaptation of the niche model of Williams & Martinez (2000). B: from an equivalent version of the Tuesday Lake ( $M, N$ )-food web re-indexed by body mass, by independently permuting the entries of each diagonal of the predation matrix (a ‘diagonal equivalence’ model). C: from an equivalent version of the Tuesday Lake ( $M, N$ )-food web re-indexed by body mass, by independently permuting the entries of each of the blocks in the predation matrix described in Appendix S1, Table A9 and Fig. A2 (a ‘block equivalence’ model). D: from an equivalent version of the Tuesday Lake ( $M, N$ )-food web re-indexed by body mass, by using a cascade-like model (‘cascade model with functional groups’). The second and third columns of the table have the statistics of the Tuesday Lake link length distributions. The subsequent columns are based on 5000 food webs produced by each model. The columns ‘min.’, ‘mean’ and ‘max.’ provide the minimum, mean and maximum of the statistics over the 5000 runs. The column ‘% less’ gives the percentage of model statistics that were less than the corresponding Tuesday Lake statistic. Columns 6 and 9 sometimes contain ‘max.’ data and sometimes contain ‘% less’ data. The ‘max.’ data are shown only when the ‘% less’ data would be 0%

	Data		Model simulations					
			Jarque–Bera			Lilliefors		
	Jarque–Bera	Lilliefors	min.	mean	% less	min.	mean	% less
<b>A</b>								
1984	6.943	0.0524	4.300	122.480	0.14%	0.0376	0.1328	0.14%
1986	3.452	0.0515	4.230	192.993	6203.35	0.0451	0.1357	0.14%
<b>B</b>								
1984	6.943	0.0524	21.648	49.149	109.115	0.0600	0.0932	0.1427
1986	3.452	0.0515	51.083	191.756	345.830	0.0603	0.1144	0.1754
<b>C</b>								
1984	6.943	0.0524	2.955	6.899	51.8%	0.0253	0.0464	79.64%
1986	3.452	0.0515	1.492	5.939	2.74%	0.0313	0.0670	5.1%
<b>D</b>								
1984	6.943	0.0524	3.651	10.488	4.04%	0.032	0.059	25.52%
1986	3.452	0.0515	1.651	6.870	1.36%	0.032	0.062	15.22%

cascade median angles were greater than the median angle of Tuesday Lake. In 1986, 19.2% of the median angles were greater than the Tuesday Lake median angle.

## Discussion

This analysis gives a new quantitative overview of the Tuesday Lake web. We summarize the most important results in two steps, first assuming that every link is identical to the average link, and then recognizing the important variation among links.

As a first approximation, suppose every link has  $l_1$  length 6 and slope  $-1$ . A slope of  $-1$  means that the biomass of predator and prey are equal, and that differences between predator and prey in average body mass and in numerical abundance contribute equally to the length of 6 orders of magnitude. Then the body mass ratio of predator to prey is 1000 (because  $\log_{10} 1000 = 3$ ), and the numerical abundance ratio of prey to predator is 1000. If body mass is proportional to body length cubed, then the typical predator is 10 times as long as its prey, while the prey is 1000 times more numerically abundant (individuals/m<sup>3</sup>) than the predator. Given that the average body mass of species in Tuesday Lake ranged over almost 12 orders of magnitude and numerical abundance varied almost 10 orders of magnitude (Cohen *et al.* 2003), the  $l_1$  distance from the smallest to the largest species was 21 or 22 orders of magnitude, a distance that could be spanned by four links each of

length 6. In fact, as Jonsson *et al.* (2004) showed, the mean length of food chains in the unlumped web was 4.6 links in 1984 and 4.2 links in 1986; in the lumped web, it was 3.7 and 3.5 links in the corresponding years.

We hypothesize that other webs in which the top species are typically the largest species in the web and basal species are typically the smallest will have mean link length equal to the community span (see Introduction) divided by the mean length of a chain in the web. These observations and predictions suggest an answer to the classic question (Hutchinson 1959) of why food chains typically have so few links: food chains are short as a result of the typical differences between predator and prey (i.e. the length of a typical trophic link is 6 orders of magnitude) in combination with the limited range of average body mass and numerical abundance over all species (here about 22 orders of magnitude). This resolution of the question replaces one question with two: why (or under what conditions) is a typical trophic link 6 orders of magnitude long, and why (or under what conditions) is the combined range of average body mass and numerical abundance limited to 22 orders of magnitude?

In 1984 and 1986, short ordered pair lengths were much more common than long ones. Any model of ( $M, N$ )-webs must select links to reproduce the observed normal link length distribution by choosing from this background distribution of pair lengths, which is heavily skewed to short pairs. Median pair lengths were between 3.4 and 4.1 orders of magnitude in both years,

while mean link lengths were about 6 in both years. The pairs that were most likely to be links had length between 9 and 11 orders of magnitude in both years. A tension between the greater likelihood that a longer pair (of length 9–11 orders of magnitude) will materialize as a link and the greater abundance of short pairs results in links whose mean length is 6 orders of magnitude (Fig. 4). The cascade and niche models produce  $(M, N)$ -webs with unrealistic link length distributions, given observed  $M$  and  $N$  distributions, because they select links of all lengths from the collection of available ordered pairs with equal likelihood.

As a second approximation, the variation in the length of links is well described by a normal distribution with mean approximately 6 and standard deviation approximately 2.7. We hypothesize that other webs will also have normal link length distributions. Normality produces a quantitative goal that models of  $(M, N)$ -webs should be required to fulfil if other  $(M, N)$ -webs confirm the pattern. Attempts to understand this normal variation yield important insights into the mechanisms that may produce it, and into the inadequacy of the models that have so far been proposed to explain other features of web structure. The block equivalence model and the cascade model with functional groups succeeded in producing normality, suggesting that a functional classification of species (phytoplankton, zooplankton, fish; or basal, herbivorous and omnivorous/carnivorous) may play an important role in producing the normality of the link length distribution. The forced link length distribution model showed that functional classification is not mathematically necessary (although it is sufficient) to explain normal link length distributions. A functional classification is probably the most biologically reasonable way to proceed.

The finding here that a typical body mass ratio between predator and prey is 1000 in the web of the pelagic Tuesday Lake community does not contradict, but differs from, earlier findings that a typical body mass ratio between competitors in a guild, such as congeneric granivorous birds, is 2 (e.g. Hutchinson 1978: 174). When such findings are based on community-wide studies and not on selected taxonomic groups or literature surveys from scattered habitats, the results provide the nucleus of a catalogue of associations between different interspecific relationships (predation, competition, parasitism, symbiosis, etc.) and body mass ratios in different types of habitats (pelagic freshwater or marine, benthic freshwater or marine, above-ground terrestrial, soil, hydrothermal, phytotelmata, etc.). Such a catalogue could increase the realism of both inputs and outputs of dynamical ecological models.

Because body mass is related allometrically to demographic parameters such as birth rate and death rate (Peters 1983), the magnitude of the coefficients of dynamic models would be constrained by species' body mass. Even in two-species Lotka–Volterra predator–prey differential equations, our results plus allometric relations constrain the turnover rates of predator and prey.

More extensive use of body mass and numerical abundance information in dynamical models could help theoretical investigations of structure and stability and management simulations in the absence of direct measurements of demographic parameters.

How general are the results obtained here for Tuesday Lake? Does either the first approximation based on averages or the second approximation recognizing variability describe the webs of other lakes? Of other aquatic or marine webs? Of terrestrial webs? Obtaining the data to answer these questions is an exciting prospect.

## Acknowledgements

R. Eugene Turner suggested the study of link lengths in conversation with J.E.C. in August 2002 at the VIIIth International Congress of Ecology in Seoul, Korea. D.C.R. thanks Tomas Jonsson for patient help in comparing data analyses. J.E.C. thanks Mr and Mrs William T. Golden for hospitality during this work. The authors thank Stephen R. Carpenter for permission to use the Tuesday Lake data, Kathe Rogerson for help, Jordi Bascompte, Carlos Melián, R. Eugene Turner, Heinrich Zu Dohna and the referees for many helpful suggestions, and the US National Science Foundation for support from grant no. DEB9981552.

## Supplementary material

The following material is available from: <http://www.blackwellpublishing.com/products/journals/suppmat/JAE/JAE856/JAE856sm.htm>

**Appendix S1.** Trophic links' lengths and slope in the Tuesday Lake food web with species' body mass and numerical abundance.

**Fig. A1.** Estimate of the 1984 frequency distribution of numerical abundance in Tuesday Lake, using a kernel density estimator.

**Fig. A2.** Predation matrices in (a) 1984 and (b) 1986.

**Fig. A3.** Trophic links and ordered pairs of species of Tuesday Lake with heavier predator than prey, plotted on  $l_1$  vs.  $\log(N_i)$ .

**Fig. A4.** Frequency distributions of lengths of all pairs and of all links in 1986.

**Table A1.** Groupings of all non-isolated species in Tuesday Lake in 1984.

**Table A2.** Groupings of all non-isolated species in Tuesday Lake in 1986.

**Table A3.** Means and SDs of  $\log(M)$  and  $\log(N)$  and the number of species in each group from the BIT- and PZF-groupings for 1984 and 1986.

**Table A4.** Comparison between fitted and theoretical equations in accounting for variance of link lengths as functions of body masses and numerical abundances.

**Table A5.** Coefficients of a linear regression of  $\alpha$  vs.  $\log(M_p)$  and  $\log(M_p)$  and  $\alpha$  vs.  $\log(N_p)$  and  $\log(N_p)$  over links and over pairs with heavier predator, where  $\alpha$  is the slope of a link (pair).

**Table A6.** Coefficients of linear best-fits and predictions of theory for the distributions ( $\log(M_p)$ ,  $l_i$ ) and ( $\log(N_p)$ ,  $l_i$ ) over links and over pairs with predator at least as heavy as prey.

**Table A7.** Results of statistical tests of the normality of distributions of angles.

**Table A8.** Results of statistical tests of the normality of distributions of lengths.

**Table A9.** Boundaries of the blocks selected in the 'block equivalence' model.

**Additional results.**

**Matlab code for models.**

## References

- Bersier, L.-F., Banasek-Richter, C. & Cattin, M.-F. (2002) Quantitative descriptors of food web matrices. *Ecology*, **83**, 2394–2407.
- Blackburn, T.M. & Gaston, K.J. (1999) The relationship between animal abundance and body size: a review of the mechanisms. *Advances in Ecological Research*, **28**, 181–210.
- Cohen, J.E. (1990) A stochastic theory of community food webs. VI. Heterogeneous alternatives to the cascade model. *Theoretical Population Biology*, **37**, 55–90.
- Cohen, J.E., Briand, F. & Newman, C.M. (1990) *Community Food Webs: Data and Theory*. Springer Verlag, Berlin, Germany.
- Cohen, J.E., Jonsson, T. & Carpenter, S. (2003) Ecological community description using the food web, species abundance, and body size. *PNAS*, **100**, 1781–1786.
- Cohen, J.E., Pimm, S.L., Yodzis, P. & Saldaña, J. (1993) Body sizes of animal predators and animal prey in food webs. *Journal of Animal Ecology*, **62**, 67–78.
- Havlicek, T. & Carpenter, S. (2001) Pelagic species size distributions in lakes: are they continuous? *Limnology and Oceanography*, **46**, 1021–1033.
- Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *American Naturalist*, **93**, 145–159.
- Hutchinson, G.E. (1978) *An Introduction to Population Ecology*. Yale University Press, New Haven.
- Jarque, C.M. & Bera, A.K. (1987) A test for normality of observations and regression residuals. *International Statistical Review*, **55**, 163–172.
- Jonsson, T., Cohen, J.E. & Carpenter, S.R. (2004) Food webs, body size and species abundance in ecological community description. *Advances in Ecological Research*, in press.
- Leaper, R. & Raffaelli, D. (1999) Defining the abundance body-size constraint space: data from a real food web. *Ecology Letters*, **2**, 191–199.
- Lilliefors, H.W. (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **62**, 399–402.
- Moore, J.C., deRuiter, P.C. & Hunt, H.W. (1993) Influence of productivity on the stability of real and model ecosystems. *Science*, **261**, 906–908.
- Peters, R.H. (1983) *The Ecological Implications of Body Size*. Cambridge University Press, New York.
- Peters, R.H. & Raelson, J.V. (1984) Relations between individual size and mammalian population density. *American Naturalist*, **124**, 498–517.
- Peters, R.H. & Wassenberg, K. (1983) The effect of body size on animal abundance. *Oecologia*, **60**, 89–96.
- Rott, A.S. & Godfray, H.C.J. (2000) The structure of a leafminer-parasitoid community. *Journal of Animal Ecology*, **69**, 274–289.
- deRuiter, P.C., Neutel, A.-M. & Moore, J.C. (1995) Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science*, **269**, 1257–1260.
- Siegel, S. (1956) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, New York.
- Ulanowicz, R.E. (1984) Community measures of marine food networks and their possible applications. *Flows of Energy and Materials in Marine Ecosystems* (ed. M.J.R. Fasham), pp. 23–47. Plenum, London.
- Williams, R.J. & Martinez, N.D. (2000) Simple rules yield complex food webs. *Nature*, **404**, 180–183.

Received 24 October 2003; accepted 23 January 2004