Current food web models cannot explain the overall topological structure of observed food webs

Jeremy W. Fox


Topological food webs illustrating “who eats whom” in different systems exhibit similar, non-random, structures suggesting that general rules govern food web structure. Current food web models correctly predict many measures of food web topology from knowledge of species richness and connectance (fraction of possible predator–prey links that actually occur), together with assumptions about the ecological rules governing “who eats whom”. However, current measures are relatively insensitive to small changes in topology. Improvement of, and discrimination among, current models requires development of new measures of food web structure. Here I examine whether current food web models (cascade, niche, and nested hierarchy models, plus a random null model) can predict a new measure of food web structure, structural stability. Structural stability complements other measures of food web topology because it is sensitive to changes in topology that other measures often miss. The cascade and null models respectively over- and underpredict structural stability for a set of 17 high-quality food webs. While the niche and nested hierarchy models provide unbiased predictions on average, their 95% confidence intervals frequently fail to include the observed data. Observed structural stabilities for all models are overdispersed compared to model predictions, and predicted and observed structural stabilities are uncorrelated, indicating that important sources of variation in structural stability are not captured by the models. Crucially, poor model performance arises because observed variation in structural stability is unrelated to variation in species richness and connectance. In contrast, almost all other measures of food web topology vary with species richness and connectance in natural webs. No model that takes species richness and connectance as the only input parameters can reproduce observed variation in structural stability. Further progress in predicting and explaining food web topology will require fundamentally new models based on different input parameters.

J. W. Fox, Dept. of Biological Sciences, Univ. of Calgary, 2500 University Dr. NW, Calgary, AB, Canada, T2N 1N4 (jefox@ucalgary.ca).

A food web is a synthetic description of an ecological community. Even topological (qualitative) food webs, which simply illustrate “who eats whom”, summarize a great deal of information on diversity, species composition, trophic structure, energy and material flows, and species interactions (Williams and Martinez 2000). For this reason ecologists have long sought patterns in food web topology (Cohen 1978, Cohen and Newman 1985). Patterns in food web topology suggest the operation of general “rules” that generate these patterns. Discovery of these rules would provide fundamental insight into the ultimate causes of community structure. For instance, if community structure ultimately reflects dynamic stability constraints, observed food web structures should be more stable than expected by chance (Yodzis 1981). There may also be surprising fundamental analogies.
between the rules governing food web structure and the rules structuring other kinds of networks, like the internet (Dunne et al. 2002). Knowledge of the rules governing food web topology also would aid progress on numerous applied issues, such as predicting the direct and indirect impacts of invasive species (Woodward and Hildrew 2001). Research on these issues will be challenging, because substantial progress likely will require detailed empirical data, and models capable of reproducing observed details (Williams and Martinez 2000, Borer et al. 2002).

Several models suggest detailed patterns in food web structure may reflect physical and/or phylogenetic constraints on “who can eat whom” (Cohen and Newman 1985, Williams and Martinez 2000, Cattin et al. 2004). For instance, in aquatic systems predators typically are larger than – but not too much larger than— their prey (Warren and Lawton 1987, Cohen et al. 1993, 2003), and in many systems phylogenetically-related species have similar diets (Cattin et al. 2004). Current models correctly predict many measures of food web topology using only two input parameters, species richness (S) and connectance (C), combined with ecological assumptions intended to capture the effects of physical/phylogenetic constraints on diet (Williams and Martinez 2000, Cattin et al. 2004). Connectance is the fraction of possible predator–prey links that are actually observed, and so equals 1 when there are $S^2$ links (i.e., every species eats every other species, including itself). The input parameters S and C together determine the number of links in the food web, while the ecological assumptions govern link arrangement. The use of S and C as input parameters is motivated in part by the observation that many patterns in food web structure take the form of simple functional relationships between measures of food web topology (dependent variables), and S and/or C (independent variables) (Cohen and Newman 1985, Williams and Martinez 2000, Cattin et al. 2004). Even those measures of food web structure that are not well-predicted by current models (e.g., numbers of “network motifs”, Bascompte and Mélián 2005) vary systematically with species richness and connectance, suggesting that slightly modified versions of current models might predict these measures as well. Indeed, the history of theoretical work on food web topology is largely a history of improving the ecological assumptions describing physical/phylogenetic constraints on link arrangement, so as to better predict various measures of food web topology (Cohen and Newman 1985, Williams and Martinez 2000, Cattin et al. 2004).

Continued progress in understanding food web topology depends on having measures of topology that, alone or in combination, are sufficiently sensitive to provide stringent tests of a model’s predictions. However, many current measures are relatively insensitive to small changes in food web structure. Current measures of food web topology lack sensitivity because they take the form of averages, sums, standard deviations, or proportions, calculated across all species or across all subparts of a web. For instance, mean food chain length is simply the average length of all linear food chains from a basal species to a top predator, and comprises an average across hundreds, thousands, or tens of thousands of food chains, depending on the complexity of the web (Williams and Martinez 2000). Randomly shuffling a few, or even many, of the links in a complex food web would be unlikely to greatly alter mean food chain length. A model could therefore correctly predict mean food chain length without actually producing a food web that closely resembles the real web.

One way around this problem is to test a model on different, complementary measures of food web structure. However, current measures of food web structure apparently are not sufficiently complementary to discriminate among similar models. The niche and nested hierarchy models produce food webs that look quite different to the naked eye (Fig. 1 in Cattin et al. 2004), but make very similar predictions for most current measures of food web topology (Cattin et al. 2004, Stouffer et al. 2005). Current food web models dramatically outperform a random null model, and so clearly capture at least some of the broad determinants of food web topology, but which models (if any) also capture the details is an open question that is difficult to answer with current measures of food web topology.

![Fig. 1. Two hypothetical food webs of 9 species, differing only in the placement of a single predator–prey link (bold arrow). Arrows point from prey to predator; looped arrows indicate cannibalism. Webs (a) and (b) are identical in many topological properties (including species richness; connectance; mean food chain length; proportions of top, intermediate, and basal species; proportion of omnivores; proportion of cannibals; mean and standard deviation of generality and vulnerability; degree distribution; see Williams and Martinez 2000 and Dunne et al. 2002 for definitions) but differ greatly in structural stability.](image-url)
Here I propose a new measure of food web topology, structural stability, and ask whether current food web models correctly predict structural stability. I first express a topological food web in matrix form, and then define structural stability as the largest real part of the eigenvalues of the food web matrix. Structural stability can be interpreted as the contribution of food web topology (which determines the signs of species interaction strengths) to local asymptotic (dynamic) stability. In contrast, many proposed measures of food web structure lack obvious ecological interpretations (e.g. the standard deviation of the lengths of all linear food chains in a web, Williams and Martinez 2000). This ecological interpretability makes structural stability more versatile than a simple abstract measure of food web structure. For instance, structural stability links topological food web theory to the study of dynamic constraints on food web structure (Yodzis 1981). Because structural stability is not a simple sum, average, standard deviation, or proportion, it is sensitive to structural changes (repositionings of predator–prey links) that do not alter other measures of food web topology such as mean food chain length (Fig. 1). Structural stability is not a perfect measure of food web structure (no single number can summarize all aspects of food web topology), but rather complements previously-proposed measures by capturing many structural changes that previous measures likely would miss.

I compared the structural stabilities of 17 observed food webs to the predictions of a random null model, and three food web models: the cascade model (Cohen and Newman 1985), the niche model (Williams and Martinez 2000), and the nested hierarchy model (Cattin et al. 2004). These three food web models differ in the details of their assumptions, but all share fundamental similarities. All three assume that the physical/phylogenetic similarity of species can be summarized with a single number (a “niche value”), and that physical/phylogenetic constraints tend to cause species with larger niche values to consume those with smaller niche values (a “trophic hierarchy”). Further, the niche and nested hierarchy models (but not the cascade model) assume that species tend not to consume those with much smaller niche values (Stouffer et al. 2005). However, all three models differ in their details (Williams and Martinez 2000, Cattin et al. 2004). Structural stability provides an opportunity to test whether a model gets the basics and the details right, since unlike many other measures of food web structure, structural stability is not analytically derivable from the shared basic assumptions of the niche and nested hierarchy models (Stouffer et al. 2005). Structural stability therefore tests the performance of current food web models in a new way, by measuring aspects of food web topology not considered previously.

Methods

Construction of qualitative webs

I considered 17 highly-detailed qualitative natural food webs: the 14 previously-published webs considered by Dunne et al. (2002), plus three additional marine webs (the Benguela system off South Africa, a small Caribbean coral reef, and a North American continental shelf) described in Dunne et al. (2004). Dunne et al. (2002) also considered two unpublished webs that I do not consider. The 17 webs come from a variety of terrestrial, freshwater, and marine communities, and previous studies describe various features of these webs (Williams and Martinez 2000, Dunne et al. 2002, 2004, Cattin et al. 2004). The webs range in richness from 25–155 trophic species (species sharing the same predators and prey). I considered trophic species so as to conform to published food web models (Williams and Martinez 2000, Cattin et al. 2004). Aggregation of biological species into trophic species reduces species richness to an average of 85% of its original value, with 12/17 webs having trophic species richness ≥90% of biological species richness.

A qualitative food web can be expressed as a matrix A where non-zero elements a_{ij} and a_{ji} (i,j = 1…S, where S is trophic species richness) indicate the existence of a trophic interaction between species i and j. I assume that a_{ij} = −1 if species j is a predator of i, and assign the corresponding element a_{ji} a value of +1. Describing the interaction between predator j and prey i with a pair of matrix elements a_{ij} and a_{ji} of differing signs, rather than with a single negative element a_{ij}, allows ecologically-meaningful calculation of structural stability. If species i and j consume one another (mutual predation), I assume a_{ij} = a_{ji} = −1, thereby treating mutual predation as a form of interference competition. I assign cannibalistic species a_{ij} = −1, thereby treating cannibalism as intra-specific interference competition. Because I focus on trophic structure, I assume no direct interactions among species that are neither prey nor predator of one another (a_{ij} = a_{ji} = 0).

Measurement of structural stability and statistical analyses

I define structural stability as the largest real part of the eigenvalues of A (max(Re(λ(A)))). In dynamical food web theory, the matrix A is interpreted as giving the interaction strengths of one species on another, and the largest real part of the eigenvalues of A measures asymptotic dynamic stability (resilience), the ability of the system to recover from a small perturbation (May 1973). In the context of dynamical food web theory, when the non-zero elements of A all have the same absolute magnitude (|a_{ij}| = 1), the elements a_{ij} are
interpreted as the signs of the interaction strengths (May 1973). May (1973) pointed out that systems with appropriate sign structures will always recover from a small perturbation, whatever the absolute magnitudes of the elements of A; such systems are called sign stable. Sign stability is an all-or-nothing property—a system is either sign stable, or not. However, when the elements of A all have the same absolute magnitude, I suggest that max(Re(λ(A))) can be interpreted as a measure of the contribution of qualitative (sign) structure to dynamic stability (for related suggestions, see Dambacher et al. 2003a, 2003b). I call this contribution structural stability. Webs with high structural stability (i.e., low max(Re(λ(A)))) when all non-zero aij have the same absolute magnitude) will be dynamically stable for many different choices of the quantitative magnitudes of the aij. Conversely, webs with low structural stability would require aij of very precisely chosen quantitative magnitudes in order to be dynamically stable (Dambacher et al. 2003a, 2003b).

Structural stability also is a useful measure of food web topology that is sensitive even to small changes in topology. This is because structural stability reflects the numbers of negative and positive feedback loops of different lengths in a food web, as well as self-feedback (non-zero diagonal elements), and feedback loops reflect the food web topology (Neutel et al. 2002, Dambacher et al. 2003b). Here, a feedback loop refers to a closed chain ofrophic effects (of predator on prey and/or prey on predator) that does not include any species more than once. Feedback loops include “loops” sensu Williams and Martinez (2000) – closed chains of effects of predators on prey – as well as other kinds of loops. Negative feedback loops are those where the product of the signs of the effects in the loop is −1 (e.g., a two-species loop from predator i to prey j) and back again has aijaij = (−1) × 1 = −1). Positive feedback loops are those where the product of the signs of the effects in the loop is +1. Negative feedback loops increase structural stability (i.e., reduce max(Re(λ(A)))) and positive feedback loops decrease stability, although structural stability also depends on the frequency of negative feedback loops of different lengths (Neutel et al. 2002, Dambacher et al. 2003b). Changing the position of even one predator–prey link in a food web will very likely change the frequency of negative and positive feedback loops of different lengths, thereby altering structural stability (Fig. 1). Structural stability of course does not capture all differences in food web topology, since it is possible for two webs of different topologies to have the same structural stability. However, I show below that this possibility is unlikely; the more two webs differ in topology, the more they are likely to differ in structural stability.

Structural stability necessarily equals zero when all or most diagonal elements aii = 0. In order to study the qualitative structure of observed and model food webs with < ~10 cannibalistic species (a category which includes 12/17 observed webs), it is therefore necessary to assume that some of the non-cannibalistic diagonal elements aii are non-zero. I assume that a randomly-selected 50% of the non-cannibalistic species in each observed web have aii = −1, which can be thought of as reflecting intraspecific competition. The remainder of the non-cannibalistic species have aii = 0. Because the value of structural stability for any web will vary slightly depending on which species have aii = −1, I evaluated the structural stabilities of 100 realizations of each observed web, each with a randomly-selected 50% of non-cannibalistic diagonal elements equal to −1. For each of the 17 webs, the mean of these 100 realizations gives the “true” observed structural stability, which a successful food web model will accurately predict. The fraction 50% necessarily is an arbitrary choice. A limited number of additional simulations indicated that results were qualitatively unchanged when 10% or 90% of non-cannibalistic species had aii = −1. Alternative treatments of the diagonal elements have serious drawbacks. The alternative, biologically-motivated choice of setting aii = −1 for all basal species (those with no prey) and aii = 0 for non-basal non-cannibals (Pimm and Lawton 1977) would make structural stability very sensitive to the proportion of basal species in the food web. This would therefore reward a model that correctly predicted the proportion of basal species while poorly predicting overall food web topology. The alternative arbitrary choice of setting all aii = −1 would be undesirable for two reasons. First, it would fail to reward models for correctly predicting the number of cannibalistic species, since aii would equal −1 whether or not species i was cannibalistic. Second, structural stability necessarily equals −1 for many low-connectance webs when all aii = −1, independent of the topology of the off-diagonal elements. For instance, seven of the 10 lowest-connectance observed webs have structural stability of −1 when all aii are set to −1. Setting all aii = −1 would therefore prevent analysis of whether food web models correctly reproduce the observed structural stabilities of low-connectance webs. Randomly setting an intermediate fraction of non-cannibalistic diagonal elements to −1 allows the topology of off-diagonal and cannibalistic links to dictate structural stability.

I analyzed three food web models (cascade, niche and nested hierarchy models), plus a random null model. Cohen and Newman (1985), Williams and Martinez (2000) and Cattin et al. (2004) describe the assumptions of the food web models. I generated null model realizations by randomly shuffling the locations of the observed effects of predators on prey, so that every null model realization had exactly the same connectance C as the corresponding observed web (note that C is always calculated without including the negative diagonal
elements randomly assigned to non-cannibalistic species). Since realized connectance can differ from that of the corresponding observed web in the other models, for these models I only used realizations having C within ±3% of the corresponding observed web, as in Williams and Martinez (2000). While it is possible to generate cascade model webs having exactly the same C as the corresponding observed webs, this is computationally slower and would not have altered the results. Each model provides the location of predator–prey links (negative effects of predator j on prey i, a_{ij} = −1). For all models, every off-diagonal element a_{ij} = −1 was assigned a corresponding a_{ji} = +1, except when a_{ij} = a_{ji} = −1, as in the observed webs.

All the models examined here are stochastic: for given values of species richness and connectance, each model can produce many different food web topologies. Each observed food web can be thought of as one realization of an unknown stochastic model, which is then assigned 100 random sets of diagonal elements for purposes of estimating its structural stability. Accordingly, comparing model predictions to the observations requires generating many realizations of the model with the same S and C as the corresponding observed web, with each model realization then assigned 100 random sets of diagonal elements so that its structural stability can be determined. I generated 200 realizations of each model corresponding to each of the 17 observed webs (17 × 200 = 3400 realizations/model × 4 models = 13 600 realizations). For each of the 13 600 model realizations, I calculated structural stability by assigning 100 random sets of diagonal elements, in exactly the same way as for the observed webs. The mean structural stability of each realization, averaged across the 100 diagonal element sets, is taken as the structural stability of that realization, just as for the observed webs. The overall mean of each set of 200 structural stabilities defines a model prediction, to which the corresponding observed structural stability is compared. The 2.5% and 97.5% points of the distribution of 200 structural stabilities define a 95% confidence interval around the model prediction, within which the structural stability of the corresponding observed web should fall 95% of the time (i.e. in ~16 out of 17 cases) if the model is successful (percentile confidence interval, Manly 1997). While more model realizations, and more sets of diagonal elements/realization, would have been desirable, calculation of eigenvalues for large food webs is computationally expensive and time-consuming. The “sampling effort” used proved adequate for addressing the questions asked.

To facilitate comparison with previous work (Williams and Martinez 2000), I also calculated normalized errors for the model predictions. Normalized error is the raw error (difference between the model prediction [overall mean of 200 realizations] and the observed value), divided by the standard deviation of the 200 realizations. An accurate model will have an overall mean normalized error (averaged across all 17 webs) near zero. If structural stabilities of a set of 200 realizations are approximately normally distributed (an assumption not made by the percentile confidence intervals), the standard deviation of these realizations defines a parametric confidence interval for the model prediction. Ninety-five percent of the 200 realizations should fall within ±1.96 standard deviations of the mean, so that an observed value within this range represents a statistically acceptable fit between prediction and observation. However, normalized errors reward models for making highly variable, imprecise predictions. To test this possibility, I calculated the standard deviation of the normalized errors for each model. The expected value of this standard deviation is 1. Larger values indicate overdispersion – the observed data are more widely dispersed around the model predictions than expected, given the variability among the model realizations. Overdispersion indicates that the model does not capture all of the sources of variability in the observed data.

I checked all model realizations for empty species (species with no predators or prey), and for identical species (species with the same predators and prey), replacing such realizations with new realizations. I lacked software to check whether any realization actually comprised two or more disconnected subwebs without containing empty species. Previous studies of other aspects of qualitative food web structure do not test for multi-species disconnected subwebs because such webs likely are very rare (R. Williams, pers. comm.). Realizations comprising disconnected multi-species subwebs are most likely to occur at low connectance, but low connectance webs also are the most likely to contain empty species and so be eliminated from the analysis. Spot checks failed to reveal any realizations comprising two or more disconnected subwebs. Note that one of the observed webs (Grassland) actually comprises disconnected multi-species subwebs (R. Williams, pers. comm.). I followed previous studies (Williams and Martinez 2000, Dunne et al. 2002) and included the Grassland web in the analyses. Dropping this web from the analyses does not qualitatively alter any of the results, and produces only minor quantitative changes (not shown).

To further explore the relationship between observed and predicted values of structural stability, I calculated the parametric correlation coefficient r between the observed structural stabilities, and the predicted values from each of the four models. This analysis asks whether variation in model predictions mirrors variation in the observed data, even if the model predictions are inaccurate and/or the observed data are overdispersed. A positive correlation suggests that the model captures...
at least some of the sources of variation in the observed data.

To better understand the source of any inaccuracy in model predictions, I tested whether both observed and predicted structural stabilities vary with S and C, the input parameters in the food web models. If most of the variation in observed structural stability is unrelated to S and C, a model that takes S and C as the only input parameters will have difficulty accurately and precisely reproducing the observed data. However, systematic variation in observed structural stability with S and C that is not accurately reproduced by current models might be reproduced by a modified model. To test whether structural stability depends on S and C in different ways in observed vs model webs, I first performed an analysis of covariance (ANCOVA) for the effects of web type (observed, cascade, niche, nested hierarchy, or null), S, C, S^2, C^2, and SC (continuous covariates), and the interactions between web type and the covariates on structural stability (n = 85 = 5 web types x 17 webs). I performed this analysis using predicted structural stabilities (overall means of 200 realizations/web), rather than using all realizations, so as to equalize sample sizes between observed and model data. However, inclusion of all model realizations in the ANCOVA did not qualitatively affect the results (not shown). I chose S, C, S^2, C^2, and SC as the covariates based on preliminary inspection of the data. The effects of several interaction terms were highly significant, indicating that the effects of the covariates were web type-dependent. To interpret this type-dependency of the effects of the covariates, I conducted separate multiple regressions of structural stability on S, C, S^2, C^2, and SC for each web type, and used a stepwise procedure to select the minimal adequate model.

To aid ecological interpretation of the results, I counted the numbers of cannibalistic links and mutual predation loops in each observed web and model realization. Variation in the frequency of cannibalism (a form of negative feedback) and mutual predation (a simple positive feedback loop) might partially explain variation in structural stability.

I constructed all food web model realizations using Microsoft Excel 2002 Visual Basic macros provided by L.-F. Bersier. I added diagonal elements for non-cannibalistic species, calculated eigenvalues, and conducted statistical analyses using R for Windows 2.1.0 (R Development Core Team 2004). Construction of all 1361 700 food web matrices (4 models x 17 food webs x 200 realizations/model/web x 100 random diagonal element sets/realization, plus 17 observed webs x 100 random diagonal element sets/web) and calculation of their eigenvalues required several hundred hours of computing time.

Testing whether structural stability sensitively measures overall food web topology

Illustrative examples like Fig. 1 suggest that structural stability generally is sensitive to small changes in food web topology, but do not provide formal evidence. Further, such examples also suggest that structural stability might be overly sensitive – structural stability might be such a complex function of food web topology that two food webs differing greatly in overall topology might be as likely to have similar structural stabilities as two topologically-similar food webs. Next I describe a test for whether the expected difference in structural stability between food webs increases monotonically as the true difference in topology between the two webs increases.

Conducting this test requires a formal measure of the “true” difference in topology between two food webs. However, defining a measure of the “true” difference in topology between two food webs is not straightforward. One plausible measure of the true difference in topology between two food webs with identical S and C is the minimum number of predator–prey links that would have to be repositioned in order to make the two webs topologically identical. I tested whether the expected difference in structural stability between webs is monotonically related to this measure of the true difference in food web topology. For each of the nine observed food webs in my dataset with S < 62, I generated partially-randomized versions with k = 1, 2, …, 20 randomly-selected links randomly repositioned (100 partially-randomized versions of each web for each value of k). Generating these data for observed webs with ≥ 62 species would have been extremely time-consuming, but there is no reason to think that results for these nine observed webs are unrepresentative. To randomly-reposition a link, I deleted the original randomly-selected link, and replaced it by creating a new link between two randomly-chosen species that were unlinked in the original observed web. I then calculated the structural stability of each partially-randomized version of each web as for the observed webs. For each observed web, the mean structural stability of the 100 partially-randomized versions with k links repositioned gives the expected structural stability of a web k links removed from the observed web. These data test whether structural stability is sensitive to small changes in food web structure, since all observed webs comprise >> 20 links. To obtain additional data on webs with many links repositioned, I used the predicted structural stabilities of webs generated by the null model. In the null model, a proportion 1-C of the randomly-placed links are expected to link predators and prey which are unlinked in the corresponding observed web. That is, webs generated by the null model have the same structure as webs in
which a fraction 1-C of the links are randomly repositioned.

Results indicate that structural stability is a good index of food web topology. For eight of nine webs, expected structural stability increases significantly with an increasing proportion of links repositioned, reaching an asymptote once a very high proportion of links are repositioned (Fig. 2). One of the webs (Grassland) has a near-random topology, so that random repositioning of links does not substantially alter its expected structural stability (Fig. 2i). The regressions remain significant for seven of eight webs if data from the null model webs are dropped from the analysis, indicating that expected structural stability is sensitive even to relatively small changes in the true food web topology. These results show that two webs with the same S and C and similar topologies are likely to have similar structural stabilities, while two webs with the same S and C but very different topologies are unlikely to have similar structural stabilities. Model predictions (each of which is the mean of 200 model realizations) should be accurate if the model tends to generate food webs that are topologically-similar to the corresponding observed webs.

However, the results shown in Fig. 2 do not entirely rule out the possibility that a food web model also could attain high accuracy while generating webs with very different topologies than those observed. This is because food web models do not generate a random sample of all the food web topologies that could be generated for given values of S and C. For instance, imagine it were the case that structural stability increases with the proportion of cannibalistic species in a web, but decreases with the proportion of omnivores (this example is purely hypothetical, and merely intended as a simple illustration of the point at hand). In this case, a food web model that generated webs with more cannibals and fewer omnivores than observed might accurately reproduce observed structural stabilities while incorrectly predicting many other features of food web topology. For this reason, structural stability provides a conservative test for structural differences between observed and model food webs; it is possible that a food web model might predict "the right structural stability for the wrong reasons". Testing food web models with other measures of food web structure provides a partial check on this possibility (Williams and Martinez 2000, Cattin et al. 2004): a model that correctly predicts many different, complementary measures of food web topology likely generates food webs that closely resemble observed webs. However, current models do not consistently reproduce observed structural stabilities, so the theoretical possibility that they might consistently predict "the right structural stability for the wrong reasons" is moot.

Note that my proposed measure of the "true" difference in food web topology cannot be used to directly compare observed food webs and webs generated

Fig. 2. Mean structural stabilities of partially-randomized versions of nine food webs (a) Bridge Brook; (b) Coachella Valley; (c) Skipwith Pond; (d) Chesapeake Bay; (e) Benguela; (f) St. Martin Island; (g) St. Marks Seagrass; (h) Small Reef; (i) Grassland, vs the proportion of links randomly-repositioned. Error bars are ±1 SD. Fitted curves are of the form $y = \frac{ax}{b + x + c}$ and when shown are statistically significant at the $z=0.05$ level.
by the cascade, niche, and nested hierarchy models, since these models assume that species are ranked along a “niche axis”. In practice, the true ranking of species in an observed food web is unknown, and so the number of link repositionings separating an observed web from a model-generated web also is unknown. Therefore, indices of food web topology are required to test food web models (Solow 2005). To my knowledge, structural stability is the first such index to be rigorously validated by comparison with a formal measure of the true difference in topology (Fig. 2).

Results

The 17 observed webs vary significantly in structural stability, much more than expected given the random variation in structural stability among the 100 different sets of random diagonal elements assigned to each web (ANOVA, $F_{16,1683} = 178.40$, $P < 0.001$). This indicates that there is substantial variation in observed structural stability, which a successful food web model should reproduce. Variation in observed structural stabilities is not simply noise resulting from random assignment of diagonal elements.

The four models examined here vary in accuracy. Overall, the cascade model tends to underestimate structural stability (i.e. predicts max(Re($\lambda(A)$)) less than that observed), while the null model tends to overestimate structural stability (Table 1). These results hold whether overall mean accuracy is measured using normalized or raw errors (Table 1). The niche and nested hierarchy models are broadly accurate on average, at least when accuracy is measured in terms of raw error (Table 1).

However, none of the four models is as accurate as would be expected for a model identical with the unknown model that generated the observed data. Ninety-five percent percentile confidence intervals for structural stability predicted by the cascade model include the observed data for only 5 of 17 webs, rather than the expected 16 of 17 (Fig. 3a). Corresponding results for the other models are: niche model, 13 of 17 webs (Fig. 3b); nested hierarchy model, 8 of 17 webs (Fig. 3c); null model, 10 of 17 webs (Fig. 3d). Similar results hold when parametric confidence intervals are

Table 1. Mean raw error (RE), mean normalized error (NE), and standard deviation of NE for each model; n=17 for each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean RE</th>
<th>Mean NE</th>
<th>Standard deviation of NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.26</td>
<td>1.02</td>
<td>1.91</td>
</tr>
<tr>
<td>Cascade</td>
<td>-0.17</td>
<td>-10.54</td>
<td>12.31</td>
</tr>
<tr>
<td>Niche</td>
<td>0.05</td>
<td>-0.02</td>
<td>1.58</td>
</tr>
<tr>
<td>Nested hierarchy</td>
<td>-0.05</td>
<td>-1.49</td>
<td>2.25</td>
</tr>
</tbody>
</table>

Fig. 3. Observed vs predicted structural stabilities for the cascade model (a), niche model (b), nested hierarchy model (c), and null model (d). Error bars are 95% percentile confidence intervals. Bold dashed lines indicate 1:1 relationships. Failure of an error bar to overlap the dashed line indicates a significant difference between predicted and observed values at the $\alpha = 0.05$ level. Filled symbols highlight two food webs (Coachella Valley and Skipwith Pond) with very similar species richness and connectance but very different observed structural stabilities.
used. For the cascade, niche, nested hierarchy, and null models, parametric 95% confidence intervals include 5 of 17 webs, 12 of 17 webs, 9 of 17 webs, and 12 of 17 webs, respectively.

The four food web models also vary in their precision. Ninety-five percent percentile confidence intervals are considerably narrower on average for the cascade model than the other models (Fig. 3), and the same is true for parametric confidence intervals (not shown). For given values of the input parameters S and C, the niche, nested hierarchy, and null models often generate 95% confidence intervals that include structural stabilities far outside the range of those observed in any natural food web (Fig. 3). This contrasts with the relatively greater precision of these models when predicting other measures of food web topology (compare Williams and Martinez 2000), and illustrates the sensitivity of structural stability to small changes in food web topology. Different realizations of the same stochastic food web model often exhibit similar values for many measures of food web topology, but very different values of structural stability (Fig. 3, Williams and Martinez 2000, Cattin et al. 2004).

Despite this lack of precision in model predictions, observed structural stabilities are actually more variable than expected from any food web model. Standard deviations of normalized errors for all for food web models are > 1, indicating overdispersion of the observed data compared to model predictions (Table 1). The especially high overdispersion for the cascade model reflects the fact that its predictions are both precise and inaccurate.

Predicted and observed structural stabilities are uncorrelated for all four models (all |r| < 0.17, all P > 0.2; Fig. 3), in contrast to the positive correlations between predicted and observed values for most previously-proposed measures of food web topology (Williams and Martinez 2000, Cattin et al. 2004). Observed and predicted structural stabilities are uncorrelated because all models predict that structural stability varies systematically with S and C, while the observed data lack such variation (Fig. 4). Removing the effect of web type and associated interactions from the ANCOVA leads to a highly significant reduction in fit (difference in residual df = 24, difference in residual SS = 3.37, F = 31.70, P < 0.001), indicating that structural stability varies with S, C, SC, S², and C² in different ways for different web types. In a multiple regression, observed structural stability does not vary significantly with S, C, SC, S², and C² (R² = 0.32, P = 0.45). In contrast, stepwise multiple regression finds that predicted structural stability varies significantly with some combination of C, S, C², and SC in all models (all R² > 0.90, all P < 0.001). Inspection of residuals indicated conformity with statistical assumptions for the ANCOVA and all multiple regressions. While P values of multiple
Regression models selected by a stepwise procedure should be interpreted cautiously (Philippi 1993), the clear conclusion is that structural stability depends in some fashion on S and C only in models, not in the observed data.

Observed webs have an average of 9.7 mutual predation loops. Null model realizations exhibit an average of >100 mutual predation loops, averaging across all realizations of all webs, while observed, niche, and nested hierarchy realizations average 6–10 mutual predations. Mutual predation is impossible in the cascade model. Variation in the number of mutual predation loops is broadly associated with variation in structural stability. Null model realizations are least-structurally stable on average (i.e. highest values of max(Re(λ(A))))

While cascade model realizations are most structurally stable, as expected if mutual predation loops (a form of positive feedback loop) generate instability. However, variation in the frequency of mutual predation loops (=number of mutual predation loops divided by the number of possible mutual predation loops) does not appear to explain variation in structural stability among observed webs, except that the Coachella Valley web is both highly-unstable and has by far the highest frequency of mutual predation loops.

On average, 15% of species in observed food webs are cannibalistic. Frequency of cannibalism is unrelated or weakly related to structural stability. Cascade model realizations lack cannibalism (a negative feedback loop), but are more structurally stable than observed webs or realizations of other models (Fig. 3, Table 1). Null model realizations have similar frequencies of cannibalism to observed webs (Williams and Martinez 2000), but are much more unstable (Fig. 3, Table 1). Among observed webs, frequency of cannibals is uncorrelated with structural stability (P = 0.21).

Discussion

Structural stability in observed and model food webs

None of the food web models examined here accurately and precisely predicts the structural stabilities of observed food webs, implying that none of these models successfully reproduces overall food web topology. The niche and nested hierarchy model are at least unbiased, but both make very imprecise predictions. For both models, the 95% confidence intervals frequently include values of structural stability far outside the observed range. This lack of precision reflects the sensitivity of structural stability to small changes in food web topology. No stochastic model will make extremely precise predictions of a highly sensitive measure of food web structure. For this reason, lack of precision per se is not a concern, so long as a model's 95% confidence intervals include the observed data 95% of the time, and so long as the observed data are not more variable than expected from the model. Unfortunately, the niche and nested hierarchy models do not satisfy either of these criteria. The 95% confidence intervals for the niche model include the observed structural stability data only 70–76% of the time, depending on whether percentile or parametric confidence intervals are used. The nested hierarchy model performs even worse in this respect. Observed structural stabilities also are overdispersed compared to the predictions of current food web models.

Even some inaccuracy and overdispersion need not indicate serious failings in a model. A model that captures some, but not all, of the sources of variation in the data would be expected to exhibit some inaccuracy and overdispersion, but its predictions would be expected to correlate positively with the observed data. However, predicted and observed structural stabilities are uncorrelated for all four models. Inaccuracy, overdispersion, and lack of correlation between observations and predictions together indicate that the underlying sources of variation in the model predictions do not match the underlying sources of variability in the observed data.

Although the best current food web models do not perform as well as might be hoped, they do outperform a random null model. Random null model predictions are highly biased, while those of the best current food web models (niche and nested hierarchy models) are not. Further, the random null model's predictions are less precise on average (broader confidence intervals) than those of the niche or nested hierarchy models. The broader confidence intervals of the random null model explain why its confidence interval coverage is as good as that of the nested hierarchy model. The equal accuracy of the random and nested hierarchy models, as measured by confidence interval coverage, should not be taken to indicate that the random null model reproduces observed food web topologies as well as the nested hierarchy model.

It is already known that current food web models, although much better than a random null model, are not perfect (Williams and Martinez 2000, Cattin et al. 2004, Bascompte and Melián 2005). However, it might be thought that by refining the ecological assumptions of current models, their predictions could be improved. After all, the most successful current models (the niche and nested hierarchy models) were developed by refining the ecological assumptions of the earlier cascade model (Williams and Martinez 2000, Cattin et al. 2004). Crucially, my results indicate that such refinement is not possible, because the cause of poor model performance is not the ecological assumptions about the nature of physical/phylogenetic constraints. Current models predict structural stability poorly because current models take species richness and connectance as the
only input parameters. All current models therefore predict (i) that two webs with similar species richness and connectance will have similar structural stabilities, and (ii) that structural stability varies systematically with species richness and connectance. But two observed webs with similar species richness and connectance can vary greatly in structural stability, and observed structural stability does not vary significantly with species richness and connectance, implying that observed structural stability cannot be predicted solely from knowledge of species richness and connectance.

The Coachella Valley and Skipwith Pond webs provide a particularly striking illustration of this point. These two webs have very similar species richness (\(S = 29\) and 25, respectively) and connectance (\(C = 0.312\) and 0.315, respectively). According to all four models, they should have similar structural stabilities (Fig. 3). However, Coachella Valley is the most structurally unstable web in the dataset, while Skipwith Pond is the third-most stable. The structural stabilities of observed food webs do not arise solely, or even mainly, from variation in species richness and connectance. In this respect, structural stability complements many other measures of food web structure, which do vary significantly with \(S\) and \(C\).

Solow (1996) and Williams and Martinez (2000) noted that several other features of food web topology (e.g. mean food chain length) are overdispersed compared to the predictions of the random, niche, and/or cascade models (see also Murtaugh and Kollath 1997, Solow 2005). However, previous work evaluates model overdispersion using low-quality food web data (Solow 1996), considers overdispersion outside the context of testing food web models (Murtaugh and Kollath 1997), or focuses on the relative overdispersion of different models (Williams and Martinez 2000). Further, previous studies of overdispersion have focused on measures of food web structure that vary with species richness and connectance in observed webs, suggesting that the overdispersion might be corrected by refining the models’ ecological assumptions. No previous work demonstrates overdispersion of observed data relative to model predictions that is not correctable by refining ecological assumptions.

Identifying covariates that explain some of the variation in structural stability not captured by current food web models might suggest new directions for model development. Unfortunately, identifying such covariates may prove difficult. For instance, it has been suggested that food web structure might vary systematically among habitats (Murtaugh and Kollath 1997). However, classifying observed webs by habitat (freshwater, terrestrial, or marine) does not explain a significant amount of variation in observed structural stability (ANOVA, \(F_{2,14} = 1.01, P = 0.39\)). Observed webs could of course be classified into a greater number of more finely-detailed habitat categories, and could be cross-classified by other characteristics (e.g. primary productivity). However, the limited number of observed food webs available would severely limit the power of any statistical test based on such detailed classification.

In summary, none of the four food web models examined here capture the main sources of variation in observed structural stability. More importantly, their shortcomings cannot be corrected without changing the identity of the input parameters. My results strongly suggest that further increasing our understanding of food web topology will require new models that use different and/or additional input parameters besides species richness and connectance (Loeuille and Loreau 2005).

**Mutual predation and cannibalism**

Variation in the frequency of cannibalism does not explain variation in structural stability. I assumed that all cannibalistic species, and half of other species, exhibited stabilizing self-damped growth, so models producing cannibalistic species more frequently should be more structurally stable, all else being equal. This is not the case, presumably because different models produce food webs that differ in many respects besides cannibalism frequency. Lack of an association between observed cannibalism frequency and structural stability may reflect the fact that most observed food webs have very few cannibals. Variation in the frequency of mutual predation loops explains some variation in structural stability, but leaves most variation unexplained. Most variation in structural stability apparently reflects more subtle aspects of food web topology. It would be interesting to examine the relative frequency of positive and negative feedback loops of various lengths in observed vs model food webs, as this should strongly affect structural stability (Neutel et al. 2002, Dambacher et al. 2003b).

**Are observed food webs dynamically stable?**

Although testing whether observed food webs are more dynamically stable than expected by chance was not the main goal of this work, structural stability does isolate one component of dynamic stability, the sign structure of the interaction matrix A (May 1973, Dambacher et al. 2003a,b). Consistent with many previous studies (Yodzis 1981, de Ruiter et al. 1995, Neutel et al. 2002, Emmerson and Raffaelli 2004), I found that observed webs are much more stable than null model webs. This shows that the conclusions of Yodzis (1981) were not artifacts of examining poor-quality food web data. My results complement previous work in that I isolated the contribution of food web topology to stability. Previous
Caveats

As with all food web studies, my conclusions depend on the quality of the data. Even recently-published webs, generated for the purpose of testing food web theory, use various lines of evidence (some less reliable than others) to document feeding links, lump together or ignore numerous biological species (especially microbes, parasites, and pathogens), and treat species with complex life histories in an ad hoc fashion (e.g. treating different life history stages of the same biological species as separate “trophic” species). While it is hard to imagine that better data would reveal observed food webs to have random structures, better data might be either more or less similar to published models than current data.

The results might also change if different assumptions were made about how to express a food web as a matrix A (Cohen et al. 1990). However, as long as the same assumptions are used to convert both observed and model food webs into matrices, it seems likely that current food web models would fail to predict observed structural stabilities.

Conclusions and future directions

Current food web models are impressive achievements. It is by no means obvious that any simple model could reproduce many of the topological features of complex food webs (Caswell 1988). However, many of the measurements of food web topology examined previously are relatively “easy targets”, in that they are relatively insensitive to small changes in food web topology. Precisely and accurately predicting more sensitive measures of food web topology is a more challenging task, one at which current food web models have little success.

Because observed variation in structural stability is unrelated to variation in species richness and connectance, any model that takes species richness and connectance as the only input parameters will not reproduce observed variation in structural stability. Efforts to improve food web models via small refinements while retaining the same fundamental assumptions regarding species richness and connectance have therefore reached a point of diminishing returns. My results raise the possibility that significant further improvements may even be impossible. Perhaps each natural food web is a special case except for the fact that, broadly speaking, species can be ranked in a “trophic hierarchy” and tend to consume species of smaller, but not too much smaller, rank (Stouffer et al. 2005).

I suggest that the way to discover if further improvement is possible is to develop fundamentally new models that sacrifice some of the simplicity of current models for increased realism. It would be particularly interesting to develop a food web model that does not require connectance to be specified as an input parameter. Treating connectance as an input parameter is common practice in other areas of network theory (Dunne et al. 2002), but seems artificial when the network in question is a food web. In food webs, connectance is a dependent rather than an independent variable which reflects the identities of the species and the factors controlling their diets (Fox and McGrady-Steed 2002). The ultimate goal of future modeling should be to not only correctly predict food web structure, but to make the right predictions for the right biological reasons. Achieving this goal likely will require sacrificing some of the simplicity of current models, but with the benefit of allowing a greater variety of powerful and informative comparisons with empirical data (including experimental data; Fox and McGrady-Steed 2002). The model of Loeuille and Loreau (2005) provides an intriguing step in this direction.
Acknowledgements – Thanks to Jennifer Dunne for providing the food web data and to Louis Bersier for providing the Excel macros to generate the model webs. The ms benefited greatly from discussion with Ed McCauley and members of the McCauley lab. Owen Petchey pointed out a statistical error in an earlier version of the ms.

References