

12 Statistical Approaches for Inferring and Predicting Food-Web Architecture

Rudolf P. Rohr, Russell E. Naisbit, Christian Mazza, and Louis-Félix Bersier

12.1 Introduction

Food webs are complex networks of trophic interactions (Cohen, 1978). Identification of the factors underlying the architecture of these networks remains a key question in ecology, in the hope that this will reveal how communities may be conserved in the face of species loss and climate change. Until now, this question has mainly been tackled with two different approaches: “a priori” (Ross, 1911), through stochastic or evolutionary models aimed at reproducing the essence of the system (Cohen and Newman, 1985; Williams and Martinez, 2000; Drossel *et al.*, 2001; Cattin *et al.*, 2004; Stouffer *et al.*, 2005; Rossberg *et al.*, 2006; Allesina *et al.*, 2008; Capitán *et al.*, 2013), and “a posteriori,” through mechanistic or statistical models aimed at inferring observed networks (Petchey *et al.*, 2008; Allesina and Pascual, 2009; Rohr *et al.*, 2010, 2016).

In the first approach, assumptions about the principles underlying food-web structure are used to construct models. Following the assumptions, model food webs are generated and compared with observed data. In almost all of these models, the input parameters are the numbers of species and the numbers of trophic links. For each species, some “abstract traits” are generated randomly (“ranks” for the cascade model, “niche values” for the niche model, etc.). The comparison with observed food webs is typically achieved indirectly, by generating a large number of networks and comparing them with observed webs through statistical descriptors, e.g., the proportion of top species, or the average chain length (Cohen *et al.*, 1990; Williams and Martinez, 2008). A limitation of this approach is that different models can yield very similar results, a well-known problem with a-priori models (e.g., Cohen, 1968).

In the second approach, observed food webs are fitted using models with the objective of identifying the underlying structure. In food-web ecology, this approach is based on mechanistic (Petchey *et al.*, 2008), or statistical (also called probabilistic) models that use biological traits (usually body size) as explanatory variables or latent traits (Rohr *et al.*, 2010; Williams *et al.*, 2010; Williams and Purves, 2011; Rohr *et al.*, 2016). Direct comparison of the inferred and observed food webs is possible, e.g., by using the percentage of correctly fitted trophic links.

The statistical approach works in general as follows. Let A_{ij} be the adjacency matrix of a food web, i.e., $A_{ij} = 1$ when prey i is eaten by predator j and $A_{ij} = 0$ otherwise. The aim of a statistical model is to infer the probability of existence of trophic links between pairs of species denoted by $P(A_{ij} = 1)$. A standard approach with such binary data is to write a model for the logit of those probabilities (Kolaczyk, 2009), which takes the following general form:

$$\text{logit}\left(P(A_{ij} = 1)\right) = \log\left(\frac{P(A_{ij} = 1)}{1 - P(A_{ij} = 1)}\right) = f(\theta, X, Z), \quad (12.1)$$

with θ the parameters, X the observed species traits as explanatory variables, and/or Z the latent traits (if present in the model). Then the likelihood of observing the network A_{ij} is given by the formula

$$L(A|\theta, X, Z) = \prod_{ij} P(A_{ij} = 1)^{A_{ij}} (1 - P(A_{ij} = 1))^{(1 - A_{ij})}. \quad (12.2)$$

Here, we first present four statistical models: the body-size, latent-traits, niche, and matching-centrality models. Second, we compare their performance in inferring empirical food webs. Third, we explain how they can be used for exploring the biological factors underlying food-web architecture. Finally, we show how the matching-centrality model can be used to infer partially observed food webs, and to predict their architecture (i.e., the trophic links that a new species would form in a network).

12.2 Body-Size Model

Our first model, the body-size model, uses species body size as explanatory variables (Rohr *et al.*, 2010). It is based on the assumption that there exists an optimal ratio between the predator and its prey body sizes. The formulation for the probability of existence of a trophic link between a predator of body-size m_j and a prey of body-size m_i is given by

$$\text{logit}\left(P(A_{ij} = 1)\right) = \alpha + \beta \log\left(\frac{m_j}{m_i}\right) + \gamma \log^2\left(\frac{m_j}{m_i}\right). \quad (12.3)$$

The parameters α , β , and γ are estimated using the maximum likelihood technique. Indeed, this model is simply a generalized linear model with a binomial distribution, and with $\log(m_j/m_i)$ and $\log^2(m_j/m_i)$ as explanatory variables. The quadratic term is used to capture the optimum in the body-size ratio. This model can also be viewed as a simplified statistical version of the allometric diet breadth model (ADBMs) of Petchey *et al.* (2008).

12.3 Latent-Traits Model

The latent-traits model is an extension of the body-size model, which aims at quantifying the structure that is left unexplained by the optimal body-size ratio. The idea is to add, for each species, parameters that quantify their behavior as prey or as predator. These parameters are called latent traits (or latent variables): they are considered as important characteristics of species, but are not measured; however, they can be estimated from the data. The model is as follows:

$$\text{logit}\left(P(A_{ij} = 1)\right) = \alpha + \beta \log\left(\frac{m_j}{m_i}\right) + \gamma \log^2\left(\frac{m_j}{m_i}\right) + v_i \delta f_j. \quad (12.4)$$

Compared to the body-size model, the additional parameters are v_i , the vulnerability traits of prey species i ; f_j , the foraging trait of predator j ; and δ , a scaling parameter proportional to the relative importance of the latent term. We constrain the scale of the latent traits as follows: $\sum_i (v_i)^2 = \sum_j (f_j)^2 = 1$. The latent term can be thought of as the dominant component of a singular value decomposition (SVD) applied to an analog of the matrix of residuals once controlling for body-size ratios, with the v_i the left (or exit) dominant singular vector, the f_j the right (or entry) dominant singular vector, and δ the dominant singular value (Hoff, 2009; Rohr *et al.*, 2010). All parameters and latent traits are fitted at the same time by maximum likelihood using a simulated annealing algorithm.

12.4 Niche Model

The probabilistic niche model (Fournier *et al.*, 2009; Williams *et al.*, 2010; Williams and Purves, 2011) aims to be a statistical counterpart of the original stochastic niche model of Williams and Martinez (2000). The original niche model assumes that each species in a food web is characterized by a niche center (v_i), a diet center (f_j), and a diet breadth (r_j), and that predators j consume all prey i whose v_i values are within an interval of size r_j , centered on f_j . In the statistical version, the niche parameters are not anymore randomly drawn, but estimated directly from the food web itself. In this sense, these parameters are latent traits. We define the model as follows:

$$\text{logit}\left(P(A_{ij} = 1)\right) = -\frac{1}{r_j} (v_i - f_j)^2 + m \quad (12.5)$$

with m the common intercept. The rationale behind the equation is the following: the closer the diet center of a predator to a niche value of a prey, the larger the probability of a trophic link; the larger the diet breadth of a predator, the higher the probability of forming a trophic link. As for the latent-traits model, the parameters (m , v_i , f_j , and r_j) are estimated by maximum likelihood using a simulated annealing algorithm. Note that top

species have no v_i terms, and basal species have no f_j and r_j terms. It is worth mentioning that this equation does not perfectly match the formulation of the original niche model, where all prey species within an interval are consumed with probability 1. This constraint is incompatible with the non-intervality of diet intervals of observed food webs (Cattin *et al.*, 2004; Bersier *et al.*, 2006; Stouffer *et al.*, 2006), and would result in a statistical model with a likelihood of zero.

12.5 Matching-Centrality Model

The matching-centrality model (Rohr *et al.*, 2016) aims to be a “generalization” of the previous models. As for the niche model, species are characterized by latent traits only. Each species as prey is described by a latent trait of centrality v_i^* , and by d latent traits of matching v_i^k ($k = 1, \dots, d$). The parameter d represents the number of matching dimensions. Similarly, each species as predator is described by a latent trait of centrality f_j^* , and by d latent traits of matching f_j^k (Rossberg *et al.*, 2006, 2010, 2013). The model is mathematically defined as follows:

$$\text{logit}\left(P(A_{ij} = 1)\right) = -\sum_{k=1}^d \lambda^k (v_i^k - f_j^k)^2 + \delta_1 v_i^* + \delta_2 f_j^* + m, \quad (12.6)$$

where the parameter $\lambda^k > 0$ is the relative importance of the matching terms of dimension k , δ_1 the relative importance of the prey centrality term, δ_2 the relative importance of the predator centrality term, and m the common intercept. The model is based on the following ideas: the smaller the difference between the matching traits of a prey and of a predator, the higher the probability that a trophic link is formed; the larger the centrality trait of a prey (of a predator), the higher the expected number of predators (of prey). As in the two previous models, the latent traits of matching and centrality are estimated by maximum likelihood using a simulated annealing algorithm. Note that top species have no v_i^* and v_i^k terms, and similarly basal species have no f_j^* and f_j^k terms; for $d > 1$, the v_i^k vectors are pairwise orthogonal, and similarly the f_j^k vectors.

12.6 Performance of the Models

We judged the performance of the models based on two criteria: the fraction of correctly fitted trophic links (Ω), and the Akaike information criterion (AIC). The fraction of correctly fitted trophic links is computed as follows: from the fitted probabilities given by the model, we construct a fitted adjacency matrix by setting to 1 the L elements corresponding to the pairs of species having the highest linking probability ($L =$ number of observed trophic interactions). Then we compute the fraction of trophic links that have been correctly fitted. This measure of performance is independent of the complexity of the model. In contrast, the equation of the AIC includes a penalty for the number of parameters of the model. In our case, the body-size model has three parameters, while all

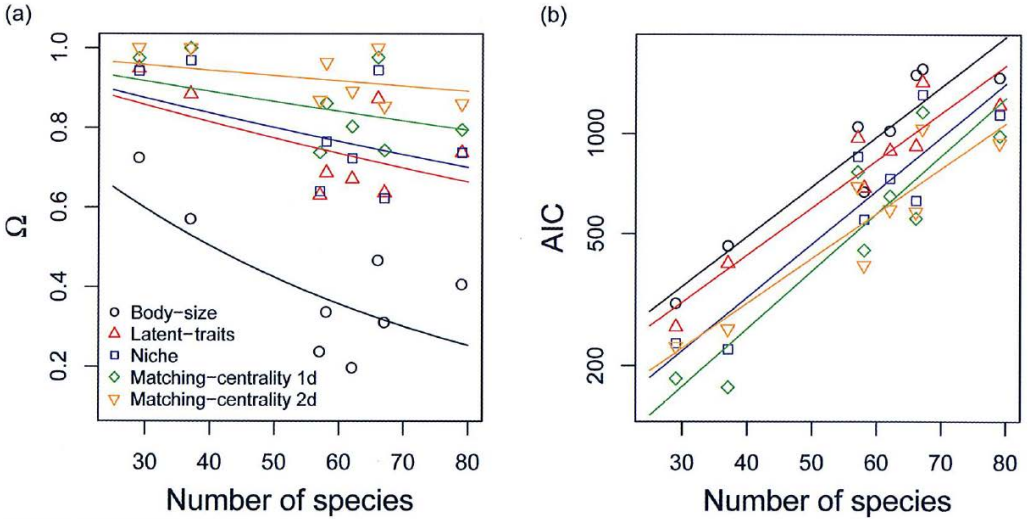


Figure 12.1 Performance of the five statistical models. Panel (a) shows the fraction of correctly fitted trophic links as a function of the number of species in the eight fitted food webs. More complex models in terms of the number of latent traits provide better fits. Panel (b) gives the AIC as a function of the number of species. It indicates that, based on this criterion, the increase in the complexity of the models is justified. (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

other models involving latent traits are much more complex, with the number of parameters scaling with the number of species in the food web.

We fitted the models to eight aquatic food webs from the published dataset of Brose *et al.* (2005): Sierra Lakes (Harper-Smith *et al.*, 2005), Tuesday Lake (Jonsson *et al.*, 2005), Mill Stream (Ledger *et al.*, unpublished data), Celtic Sea (Pinnegar *et al.*, 2003), Mulgrave River (Rayner, unpublished data), Skipwith Pond (Warren, 1989), Sheffield (Warren, unpublished data), and Broadstone Stream (Woodward *et al.*, 2005); the number of species in these systems vary from 29 to 79.

The results of fitting the body-size model, latent-traits model, niche model, and matching-centrality model in one and two dimensions are shown in Figure 12.1. Regarding the fraction of correctly fitted trophic links, we expectedly find that more complex models do perform better (Figure 12.1a). Expectedly, the decline in performance with food-web size is less pronounced for more complex models, with the two-dimensional matching-centrality being the less affected. Comparison of the models based on AIC provides the same ranking (Figure 12.1b). This is an interesting result given the large number of parameters of the models involving latent traits compared to the body-size model. It indicates that the improvement in the goodness-of-fit is not a mere consequence of increasing the complexity. Again, the matching-centrality model largely outperforms the other ones. In food webs with less than 60 species, two matching dimensions are superfluous, but provide a clear improvement for larger food webs. It is likely that more dimensions would be needed for very large webs.

12.7 Linking Latent Traits to Biological Information

Figure 12.2 represents the food web of Tuesday Lake fitted with the different models. The dots are the observed trophic interactions ($A_{ij} = 1$); the species in their role of prey (rows) and of predator (columns) are ordered according to body size or to the different

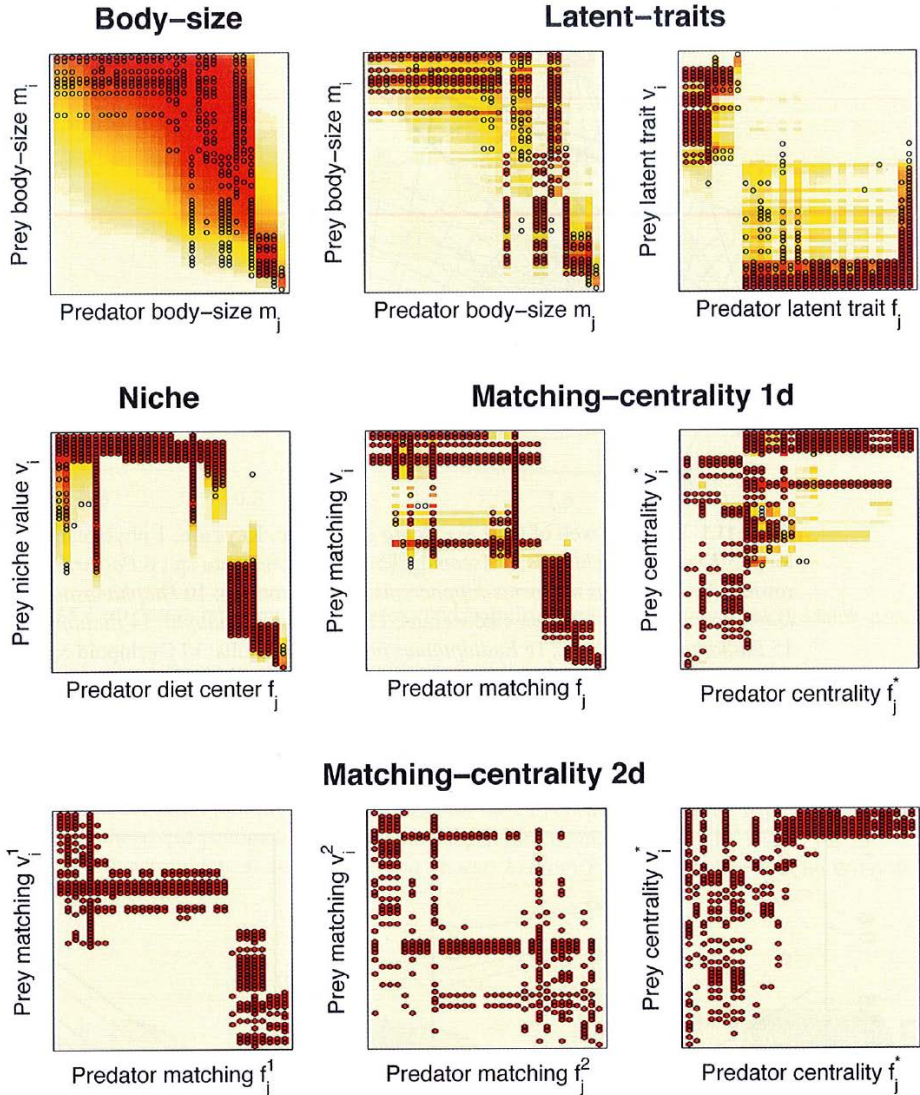


Figure 12.2 Representation of the food web of Tuesday Lake in the latent-traits space. Each panel represents the adjacency matrix, with the dots indicating a trophic interaction between a predator (columns) and a prey (row). The color, from yellow to red, indicates increasing fitted linking probability of the respective models. Species are ordered according to the relevant variable or latent trait (see axis legends). (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

latent traits; the background color gives the fitted probability that a trophic link exists between species i and j . Once the latent traits are estimated for each species (latent traits of vulnerability and of foraging in the latent-traits model; niche values, diet center, and diet breadth in the niche model; matching and centrality traits of vulnerability and of foraging in the matching-centrality model), it can be tested if they are related to biological information about the species. This can be achieved with different statistical approaches depending on the type of biological information available (e.g., simple non-parametric correlation with a quantitative variable describing the species).

In our case, information on body size and on taxonomy (as a proxy for phylogeny) is available for the eight food webs, but we present the results only for the food web of Tuesday Lake. Because taxonomic information is a qualitative variable, we choose partial Mantel tests (Legendre and Legendre, 1998) for this analysis; in this way, we can compare the relative importance of body size and of taxonomy in explaining the latent traits (Naisbit *et al.*, 2012). We compute distance matrices for the biological variables and the relevant latent traits. The distance in body size between two species i and j is computed as the absolute value in log differences, i.e., $d_{ij}^{BS} = \text{abs}(\log(m_i) - \log(m_j))$. For the distance in phylogeny, we used as a proxy the proportional number of taxonomic levels for which two species differ (Naisbit *et al.*, 2012). The distance in latent traits is simply given by the absolute value of the differences (e.g., for the latent traits of vulnerability, $d_{ij} = \text{abs}(v_i - v_j)$). We relate these three distance matrices using partial Mantel tests with Pearson correlation.

For the latent-traits model, we find that latent traits of vulnerability and foraging are both strongly related to the phylogeny (Table 12.1). This indicates that phylogeny has an effect on network structure that is independent of the optimal body-size ratio between the predators and prey. Interestingly, body size of the species in their role of prey is still correlated with the latent trait of vulnerability (once phylogeny has been accounted for); thus the optimal ratio does not fully capture the effect of body size on network structure for the species in the role of prey (Rohr *et al.*, 2010). The reason underlying this result is difficult to unravel; one possibility is that the distribution of body-mass ratios is strongly skewed, generating a correlation between body mass and latent traits. However, why this effect is apparent only for species in their role of prey requires further investigation.

The results are consistent for the other models based on latent traits only. In the niche model, the niche trait is related to the position of the species on an abstract axis depicting their position as prey, while the diet center depicts their position as predator. In the matching-centrality model, this corresponds to the vulnerability and to the foraging traits, respectively. We find that body size is correlated to all these latent traits once phylogeny is accounted for (since we use partial Mantel tests; see Table 12.1). Globally, this observation indicates that species of similar body size tend to have similar trophic roles, which is expected in aquatic systems (e.g., Cohen *et al.*, 2003).

Another important result is the presence of a strong phylogenetic signal in almost all latent traits, which indicates that phylogenetic constraints are an important determinant of the trophic structure of communities (Cattin *et al.*, 2004; Naisbit *et al.*, 2012). Thus taxonomically similar prey species tend to be eaten by taxonomically similar predators,

Table 12.1 Results of partial Mantel tests between the latent traits and body size and phylogeny for the food web of Tuesday Lake.

Type of latent trait	Body size		Phylogeny	
	partial-r	p-value	partial-r	p-value
Latent-traits model				
Latent traits of vulnerability	0.545	0.001	0.272	0.001
Latent traits of foraging	-0.010	0.472	0.341	0.003
Niche model				
Niche traits	0.785	0.001	0.192	0.001
Diet center	0.714	0.001	-0.070	0.720
Diet range	0.074	0.205	-0.056	0.772
Matching-centrality 1D model				
Vulnerability matching	0.590	0.001	0.267	0.001
Foraging matching	0.810	0.001	-0.014	0.550
Vulnerability centrality	0.294	0.004	0.074	0.031
Foraging centrality	0.649	0.001	0.064	0.183
Matching-centrality 2D model				
Vulnerability matching 1st dim.	0.330	0.001	0.070	0.010
Foraging matching 1st dim.	0.221	0.040	0.149	0.031
Vulnerability matching 2nd dim.	0.194	0.003	0.267	0.001
Foraging matching 2nd dim.	0.255	0.014	0.402	0.001
Vulnerability centrality	0.075	0.161	0.098	0.002
Foraging centrality	0.058	0.203	0.522	0.001

Legend: dim. = dimension.

Bold values indicate significant correlation at the $p \leq 0.05$ level.

and similar predator species tend to consume similar prey. This result can be seen as unsurprising, as phylogeny is a powerful integrator of the ecological characteristics of species (body size included). However, since we use partial Mantel tests, this result indicates that ecological traits uncorrelated to body size are also necessary to account for the structure of trophic interactions of prey species (e.g., in terrestrial systems, similar secondary compounds in phylogenetically related plant species; Price, 2003).

12.8 Reconstruction of Partially Observed Food Webs

One of the advantages of statistical models for food-web architecture is their ability of inferring and reconstructing partially observed food webs. In such a case, no information is available for the existence of a link for a fraction of species pairs (for those species pairs, $A_{ij} = \text{NA}$). A statistical model can be fitted on a partially observed adjacency matrix: the likelihood function (Eq. 12.2) is simply computed only on the part that is known. Mathematically this likelihood L^P is given by

$$L^P(A|\theta, X, Z) = \prod_{A_{ij} \neq \text{NA}} P(A_{ij} = 1)^{A_{ij}} (1 - P(A_{ij} = 1))^{(1-A_{ij})} \quad (12.7)$$

The inference of all the parameters and all the latent traits can be achieved with only the known part of the adjacency matrix. Using the equation of the models (Eqs. 12.3 to 12.6), we can estimate the linking probabilities for all pairs of species, in particular for the pairs of species set to NA. With these estimated linking probabilities, we can finally reconstruct the food web by predicting the presence and absence of trophic links for all pairs of species. The prediction works as follows: first, the estimated total number of trophic interactions L_P is computed as the sum over all species pairs of the linking probabilities ($L_P = \sum_{ij} P(A_{ij} = 1)$); second, the L_P pairs with the highest linking probability are set to 1 in the adjacency matrix.

We test the reconstruction of the Tuesday Lake food web using the matching-centrality model with two dimensions of matching. We simulate partially observed networks by setting to NA a given fraction of the elements of the adjacency matrix. Then, we reconstruct these generated partially observed food webs and compare the outputs with the observed matrix. As a measure of the performance of the method, we use the fraction of 1s and 0s correctly predicted. Figure 12.3a shows the results for 10% and 30% of the matrix elements set to NA. In general, the ability of the model to reconstruct Tuesday Lake's food web is quite high. Note, however, that these values have to be considered against the baseline given by the fraction of non-trophic links (one minus the connectance; the horizontal dashed line in Figure 12.3): trivially predicting an absence of trophic link for all pairs of species would result in a fraction of correct predictions equal to the fraction of non-trophic links.

12.9 Forecasting Trophic Interactions

A very interesting feature of statistical models for food-web architecture is the possibility to forecast the trophic links that a new incoming species will make when joining an existing community. For such a forecast to be sensible, it is first necessary to have a model able to faithfully infer the network. For this reason, we choose the 2D matching-centrality model. The core of the methodology resides in the use of latent traits as intermediate between the linking probabilities and observed species biological traits. We explain the method using body size and species phylogeny as biological information. The procedure is as follows: the first step consists of fitting the matching-centrality model to the observed food web. From the result, we extract the estimated latent traits of matching and of centrality for each species. The second step consists of relating each matching and centrality trait of all species to the available biological traits. In our case, this is achieved by a phylogenetic regression (Grafen, 1989; Freckleton *et al.*, 2002): we assume each latent trait to be linearly related to the log of the body size \bar{m} , and the phylogeny to induce a correlation structure. This linear model (here for a matching trait of vulnerability) is mathematically given by

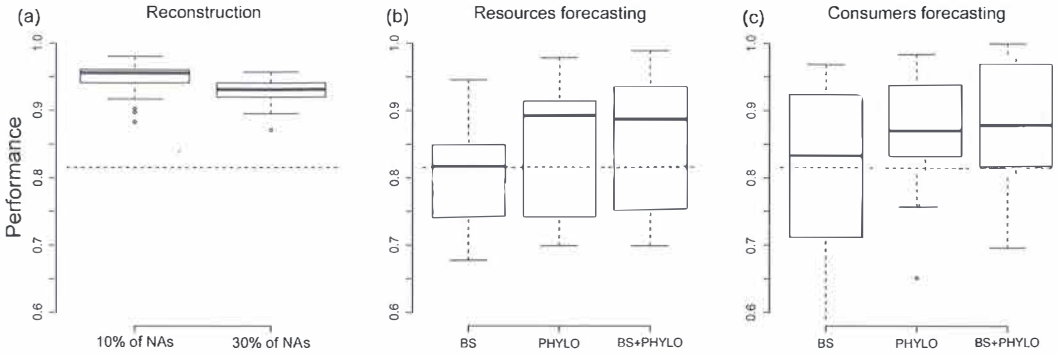


Figure 12.3 Performance of the 2D matching-centrality model in reconstructing and forecasting the Tuesday Lake food web. In all panels, the y axis gives the proportion of correctly predicted presence and absence of trophic links (0s and 1s in the adjacency matrix); the dashed lines give the baseline performance given by the proportion of 0s in the adjacency matrix (see text). Panel (a) shows the performance of reconstructing the food web after having set to NA (absence of information) a given percentage of the adjacency matrix. Panels (b) and (c) give the performance of forecasting the set of prey and the set of predators, respectively. We perform an out-of-sample test by removing triplets of species.

$$\vec{v} \sim N\left(\alpha + \beta \cdot \log(\vec{m}), \Sigma(\lambda)\right) \tag{12.8}$$

with α and β the intercept and slope, respectively, and $\Sigma(\lambda)$ the variance–covariance matrix induced by the phylogenetic relatedness. We use Pagel’s λ correlation structure (Freckleton *et al.*, 2002): the elements of the variance–covariance matrix are then given by

$$\sum (r)_{ij} = \sigma^2 \begin{cases} 1 & \text{if } i = j \\ \lambda t_{ij} & \text{if } i \neq j \end{cases} \tag{12.9}$$

where t_{ij} is the fraction of common time between species i and j on the phylogenetic tree, λ quantifies the strength of the correlation induced by the phylogeny, and σ^2 is the common variance. The third step consists of forecasting the latent traits of matching and of centrality of the new species k . Using the estimated parameters of the phylogenetic regressions of step two, we can use the conditional expectation to forecast these values, which are given by

$$\hat{v}_k = \hat{\alpha} + \hat{\beta} \cdot \log(m_k) + \sum (\hat{r})_{-k,k} \sum (\hat{r})_{k,k}^{-1} (\vec{v}_{-k} - \hat{v}_{-k}) \tag{12.10}$$

where $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ are the fitted parameters from the second step; $\sum (\hat{r})_{-k,k}$ is the k th column without the k th row (indicated by subscript $-k$) of the variance–covariance matrix; $\sum (\hat{r})_{k,k}$ is the (k,k) element of the variance–covariance matrix; $(\vec{v}_{-k} - \hat{v}_{-k})$ is the row vector of residuals obtained from the phylogenetic regression of the second step. The final step consists of estimating the linking probabilities for the new species and then

forecasting the presence or absence of trophic links. Based on the forecasted matching and centrality traits and using Eq. (12.6), we forecast the linking probabilities between the new species and the species already present; then using the same technique as in the reconstruction method, we forecast the presence or absence of trophic links.

To test our forecasting method, we perform an out-of-sample test by removing three species at the same time and then trying to forecast their trophic links in the food web of Tuesday Lake. As a measure of performance, we use the fraction of correctly predicted elements of the adjacency matrix for each triplet of test species. Figures 12.3b and c show that using body size alone as biological information provides only poor predictions; phylogeny is necessary to attain good performance. With this information, the model performs in general very well in forecasting trophic links for species, both in their role of prey and of predator.

12.10 Discussion

The motivation behind the recent development of statistical models for food-web structure was to offer a simple and intuitive tool to explore the factors underlying the architecture of ecological networks. They represent a complement to stochastic models, such as the cascade model (Cohen and Newman, 1985) and its successors (Williams and Martinez, 2000; Cattin *et al.*, 2004; Stouffer *et al.*, 2005; Rossberg *et al.*, 2006; Allesina *et al.*, 2008; Capitán *et al.*, 2013), and are more versatile than the ADBM of Petchev *et al.* (2008), which was developed specifically to explore how allometry and optimal foraging can be used to infer trophic interactions. A problem with stochastic models, which generate families of networks intended to reproduce the structure of real food webs, is that different assumptions used to build the models can lead to very similar results, a problem well-known for example with species-abundance models (Cohen, 1968). Statistical models can provide a more direct assessment of the relationships between putative underlying factors and network structure. Here we use body size and phylogeny as factors, but it would be very interesting to include other ecological variables in the models. Such information is becoming available for food webs, for example in Eklöf *et al.* (2013).

Another key advantage of statistical models is that they can be used to reconstruct partially observed systems, or to forecast the links produced by a new species. Any statistical model could theoretically be used for these purposes. For example, a version of the latent-trait model has been used to infer the linking probabilities for a plant–herbivore network (Pellissier *et al.*, 2013). Here we base our analyses on the matching-centrality model (Rohr *et al.*, 2016) since it offers by far the best fitting capabilities. We believe that these features could be very useful, for example to build so-called metawebs (Gravel *et al.*, 2013) from partial information on a system, or to estimate the trophic role of an introduced species. The approach also has its limits. For the reconstruction of food webs, the predicted trophic interactions should of course not be taken for granted, but could serve as a guide to direct sampling effort on the system in a cost-efficient way. For forecasting the links that a new species would create, the

approach is meaningful only if the new species shares enough ecological characteristics with some of the species present in the system.

The first model including latent variables in food-web ecology is the latent-trait model (Rohr *et al.*, 2010). This model requires the estimation of a large number of parameters, which is achieved with Morkov Chain Monte Carlo (aka MCMC) or simulated annealing methods (as in the present contribution). As with the other models based on latent variables, the goodness-of-fit is impressive, which may not seem surprising given the large number of parameters. However, one important outcome of our analyses is that these parameters are not merely abstract values, but are bearing meaningful biological information. Another key aspect is the possibility to formulate much more complex (and hopefully sensible) models than classical generalized linear models (e.g., the body-size model). The probabilistic-niche model (Fournier *et al.*, 2009; Williams *et al.*, 2010; Williams and Purves, 2011) is an excellent example: it is possible from the food-web matrix to infer the parameters of the classical niche model. The niche model takes the point of view of the predators to constrain the possible prey that enter their diet (Williams and Martinez, 2000); however, there is no constraint for the species in their role of prey. The matching-centrality model was developed to circumvent this limitation, with a formulation that is symmetric for prey and predators. The first step was to separate the species in their role of prey and in their role of predator in two sets. In doing so, the food web is expressed as a bipartite network, with the intermediate species appearing in both sets. The traits of vulnerability and of foraging are estimated for intermediate species, while basal species have no foraging traits, and similarly top species have no vulnerability traits. The main difference with the niche model is that prey species have a centrality trait in the matching-centrality model, which is related to the number of species they prey upon (the centrality trait for the predators is akin to the range of the niche model). Interestingly, it appeared that the simplest formulation accounting for these desiderata yielded an equation similar to that found in Rossberg *et al.* (2010), from which we named the latent traits of our model. We found out that the matching-centrality model is very versatile in decomposing any adjacency matrix into several quantitative traits for the nodes, from which the adjacency matrix can then be reconstructed. We applied this model to the analysis, reconstructing and forecasting networks as diverse as social, terrorist-association, aggression between countries, genetic, and neuronal (Rohr *et al.*, 2016). We hope that this model that emanates from food-web ecology will be useful in a wide range of scientific domains.

Acknowledgments

This work was supported by the National Centre of Competence in Research “Plant Survival” to LFB, by SystemsX.ch, the Swiss Initiative in Systems Biology to LFB and CM, and by the Swiss National Science Foundation grant 31003A_138489 to LFB.

References

- Allesina, S. and Pascual, M. (2009). Food web models: a plea for groups. *Ecology Letters*, **12**, 652–662.
- Allesina, S., Alonso, D., and Pascual, M. (2008). A general model for food web structure. *Science*, **320**, 658–661.
- Bersier, L. F., Cattin, M. F., Banasek-Richter, C., Baltensperger, R., and Gabriel, J. P. (2006). Box B: Reply to Martinez and Cushing. In *Ecological Networks: Linking Structure to Dynamics in Food Webs*, ed. M. Pascual and J. A. Dunne, New York: Oxford University Press, pp. 91–92.
- Brose, U., Cushing, L., Berlow, E. L., *et al.* (2005). Body sizes of consumers and their resources. *Ecology*, **86**, 2545.
- Capitán, J. A., Arenas, A., and Guimerà, R. (2013). Degree of intervality of food webs: from body-size data to models. *Journal of Theoretical Biology*, **334**, 35–44.
- Cattin, M. F., Bersier, L. F., Banasek-Richter, C., Baltensperger, R., and Gabriel, J. P. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature*, **427**, 835–839.
- Cohen, J. E. (1968). Alternate derivations of a species-abundance relation. *American Naturalist*, **102**, 165–172.
- Cohen, J. E. (1978). *Food Webs and Niche Space*. Princeton, NJ: Princeton University Press.
- Cohen, J. E. and Newman, C. M. (1985). A stochastic theory of community food webs. 1. Models and aggregated data. *Proceedings of the Royal Society B: Biological Sciences*, **224**, 421–448.
- Cohen, J. E., Briand, F., and Newman, C. M. (1990). *Community Food Webs, Data and Theory*. Berlin: Springer-Verlag.
- Cohen, J. E., Jonsson, T., and Carpenter, S. R. (2003). Ecological community description using the food web, species abundance, and body size. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1781–1786.
- Drossel, B., Higgs, P. G., and McKane, A. J. (2001). The influence of predator–prey population dynamics on the long-term evolution of food web structure. *Journal of Theoretical Biology*, **208**, 91–107.
- Eklöf, A., Jacob, U., Kopp, J., *et al.* (2013). The dimensionality of ecological networks. *Ecology Letters*, **16**, 577–583.
- Fournier, T., Rohr, R. P., Scherer, H., Mazza, C., and Bersier, L. F. (2009). How to estimate the niche values in a food web? The 94th ESA Annual Meeting (August 2–7, 2009), abstract: <http://eco.confex.com/eco/2009/techprogram/P20894.HTM>.
- Freckleton, R. P., Harvey, P. H., and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist*, **160**, 712–726.
- Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **326**, 119–157.
- Gravel, D., Poisot, T., Albouy, C., Velez, L., and Mouillot, D. (2013). Inferring food web structure from predator–prey body size relationships. *Methods in Ecology and Evolution*, **4**, 1083–1090.
- Harper-Smith, S., Berlow, E. L., Knapp, R. A., Williams, R. J., and Martinez, N. D. (2005). Communicating ecology through food webs: visualizing and quantifying the effects of stocking alpine lakes with trout. In *Dynamic Food Webs: Multispecies*

- Assemblages, Ecosystem Development, and Environmental Change*, ed. P. C. de Ruiter, V. Wolters, and J. C. Moore, Elsevier Academic Press, pp. 407–423.
- Hoff, P. D. (2009). Multiplicative latent factors models for description and prediction of social networks. *Computational and Mathematical Organization*, **15**, 261–272.
- Jonsson, T., Cohen, J. E., and Carpenter, S. R. (2005). Food webs, body size, and species abundance in ecological community description. In *Advances in Ecological Research*, Vol. 36, ed. H. Caswell, Elsevier Academic Press, pp. 1–84.
- Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*. New York: Springer.
- Legendre, P. and Legendre, L. (1998). *Numerical Ecology*. Amsterdam: Elsevier.
- Naisbit, R. E., Rohr, R. P., Rossberg, A. G., Kehrl, P., and Bersier, L.-F. (2012). Phylogeny versus body size as determinants of food web structure. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3291–3297.
- Pellissier, L., Rohr, R. P., Ndiribe, C., et al. (2013). Combining food web and species distribution models for improved community projections. *Ecology and Evolution*, **3**, 4572–4583.
- Pétchey, O. L., Beckerman, A. P., Riede, J. O., and Warren, P. H. (2008). Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 4191–4196.
- Pinnegar, J. K., Trenkel, V. M., Tidd, A. N., Dawson, W. A., and Du Buit, M. H. (2003). Does diet in Celtic Sea fishes reflect prey availability? In *Annual Symposium of the Fisheries Society of the British Isles*, Norwich, pp. 197–212.
- Price, P. W. (2003). *Macroevolutionary Theory on Macroecological Patterns*. Cambridge, UK: Cambridge University Press.
- Rohr, R. P., Scherer, H., Kehrl, P., Mazza, C., and Bersier, L. F. (2010). Modeling food webs: exploring unexplained structure using latent traits. *American Naturalist*, **173**, 170–177.
- Rohr, R. P., Naisbit, R. E., Mazza, C., and Bersier, L. F. (2016). Matching-centrality decomposition and the forecasting of new links in networks. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20152702.
- Ross, R. (1911). Some quantitative studies in epidemiology. *Nature*, **87**, 466–467.
- Rossberg, A. G. (2013). *Food Webs and Biodiversity: Foundations, Models, Data*. Wiley.
- Rossberg, A. G., Matsuda, H., Amemiya, T., and Itoh, K. (2006). Food webs: experts consuming families of experts. *Journal of Theoretical Biology*, **241**, 552–563.
- Rossberg, A. G., Brannstrom, A., and Dieckmann, U. (2010). How trophic interaction strength depends on traits. *Theoretical Ecology*, **3**, 13–24.
- Stouffer, D. B., Camacho, J., Guimera, R., Ng, C. A., and Nunes Amaral, L. A. (2005). Quantitative patterns in the structure of model and empirical food webs. *Ecology*, **86**, 1301–1311.
- Stouffer, D. B., Camacho, J., and Amaral, L. A. N. (2006). A robust measure of food web intervality. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19015–19020.
- Warren, P. H. (1989). Spatial and temporal variation in the structure of a freshwater food web. *Oikos*, **55**, 299–311.
- Williams, R. J. and Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, **404**, 180–183.

- Williams, R. J. and Martinez, N. D. (2008). Success and its limits among structural models of complex food webs. *Journal of Animal Ecology*, **77**, 512–519.
- Williams, R. J. and Purves, D. W. (2011). The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs. *Ecology*, **92**, 1849–1857.
- Williams, R. J., Anandanadesan, A., and Purves, D. (2010). The probabilistic niche model reveals the niche structure and role of body size in a complex food web. *PLoS ONE*, **5**, e12092.
- Woodward, G., Speirs, D. C., and Hildrew, A. G. (2005). Quantification and resolution of a complex, size-structured food web. In *Advances in Ecological Research*, Vol. 36, ed. H. Caswell, Elsevier Academic Press, pp. 85–135.