Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 241 (2006) 552–563
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Food webs: Experts consuming families of experts

A.G. Rossberg*, H. Matsuda, T. Amemiya, K. Itoh

Yokohama National University, Graduate School of Environment and Information Sciences, Yokohama 240-8501, Japan

Received 11 November 2005; received in revised form 21 December 2005; accepted 24 December 2005
Available online 7 February 2006

Abstract

Food webs of habitats as diverse as lakes or desert valleys are known to exhibit common "food-web patterns", but the detailed mechanisms generating these structures have remained unclear. By employing a stochastic, dynamical model, we show that many aspects of the structure of predatory food webs can be understood as the traces of an evolutionary history where newly evolving species avoid direct competition with their relatives. The tendency to avoid sharing natural enemies (apparent competition) with related species is considerably weaker. Thus, "experts consuming families of experts" can be identified as the main underlying food-web pattern. We report the results of a systematic, quantitative model validation showing that the model is surprisingly accurate.

Keywords: Food webs; Network structure; Competition; Evolution; Parasites

1. Introduction

The question what determines the structure of natural food webs has been listed among the nine most important unanswered questions in ecology (May, 1999). It arises naturally from many problems related to ecosystem stability and resilience (McCann, 2000; Yodzis, 1998). The difficulty of this question stems not only directly from the complexity of ecological interaction networks, but also from technical problems with recording, interpreting, and modelling food webs (Cohen et al., 1993a). For example, habitats are often not delineated clearly enough to define sharply which species to include in a web (Thompson and Townsend, 2005). The commonly used concept of binary (yes/no) trophic links is problematic, because it turns out that by various measures (Berlow et al., 2004) weak links are more frequent than strong links in natural food webs, and network structures depend on a somewhat arbitrary thresholding among the weak links (Bersier et al., 1999; Goldwasser and Roughgarden, 1993; Martinez et al., 1999; Wilhelm, 2003). Furthermore, the large number of species interacting in ecosystems has forced researchers recording food webs to disregard whole subsystems or to coarsen the taxonomic resolution (Briand and Cohen, 1984). Despite these and many other complications, statistical comparisons have revealed several regularities, *food-web patterns*, among independently recorded food webs of habitats as diverse as Caribbean islands, deserts, and lakes (e.g., Camacho et al., 2002; Cattin et al., 2004; Garlaschelli et al., 2003; Milo et al., 2002; Neutel et al., 2002; Williams and Martinez, 2000)—indicating that some simple, robust mechanism structuring food webs is at work. But the precise nature of this mechanism has remained unclear.

Evolutionary dynamics (Amaral and Meyer, 1999; Cattin et al., 2004) and pattern selection by population-dynamical stability (Yodzis, 1981) have been suggested as factors determining food-web structure. Several dynamical models containing both ingredients (Caldarelli et al., 1998; Drossel et al., 2001; Ito and Ikegami, 2006; Loeuille and Loreau, 2005; Tokita and Yasutomi, 2003; Yoshida, 2003) have been investigated. However, the large range of relevant time scales involved in these models makes their evaluation difficult. It has remained unclear which effects are essential for generating the known patterns and if these models reproduce empirical data similarly well as statistically validated, descriptive food-web models (Cattin et al., 2004; Cohen et al., 1990; Stouffer et al., 2005; Williams and

*Corresponding author. Tel.: +81 45 339 4369; fax: +81 45 339 4353.
E-mail addresses: rossberg@ynu.ac.jp (A.G. Rossberg), matsuda2@ynu.ac.jp (H. Matsuda), amemiyat@ynu.ac.jp (T. Amemiya), itohkimi@ynu.ac.jp (K. Itoh).

0022-5193/$ - see front matter
doi:10.1016/j.jtbi.2005.12.021

Martinez, 2000) such as the niche model (Williams and Martinez, 2000) or the nested hierarchy model (Cattin et al., 2004).

Here we argue that the main mechanism structuring food webs is the evolution of the pool of species adapted to the habitat[1] considered, which leads to homologies between related species in the traits determining the vulnerabilities of species as resources (prey) and in the traits determining foraging strategies and capabilities. Using a stochastic dynamical model, we show that food-web structures as empirically observed can be obtained by such a mechanism. Our analysis takes the difficulties cited above into account by employing a quantitative link-strength concept, an appropriate data standardization, a rigorous model validation procedure, and by reflecting the inhomogeneity of empirical methodology in our food-web model and data analysis.

The model proposed here ("matching model") has some similarity with the "speciation model" that we had investigated earlier (Rossberg et al., 2005,2006). But the speciation model lacked crucial elements of the matching model, e.g., the trait matching between consumers and resources. As a result, it could reproduce some features of empirical food-webs, such as vulnerability or generality distributions (Camacho et al., 2002; Stouffer et al., 2005) only under unrealistic assumptions regarding the allometric scaling of evolution rates (Rossberg et al., 2006). There had also been difficulties with reproducing empirical food webs when their size exceed some 50 or so species. These problems seem to be overcome by the matching model. The current work also considerably improves the model validation procedure, thus allowing for the first time a comparison between food-web models based on the Akaike Information Criterion.

2. The model

The matching model describes the evolution of an abstract species pool. For each species in the pool the traits determining its foraging strategies and capabilities and the traits determining its vulnerability to foraging (Caldarelli et al., 1998; Drossel et al., 2001; Yoshida, 2003) are modeled by two sequence of ones and zeros of length n (the reader might think of oppositions such as sessile/vagile, nocturnal/diurnal, or benthic/pelagic). The strength of trophic links increases (nonlinearly) with the number m of foraging traits of the consumer that match the corresponding vulnerability traits of the resource (Fig. 1). A trophic link is considered as present if the number of matched traits m exceeds some threshold $m \geqslant m_0$. In addition, each species is associated with a size parameter s characterizing the (logarithmic) body size of a species ($0 \leqslant s < 1$). Consumers cannot forage on species with size parameters larger than their own by more than λ. The

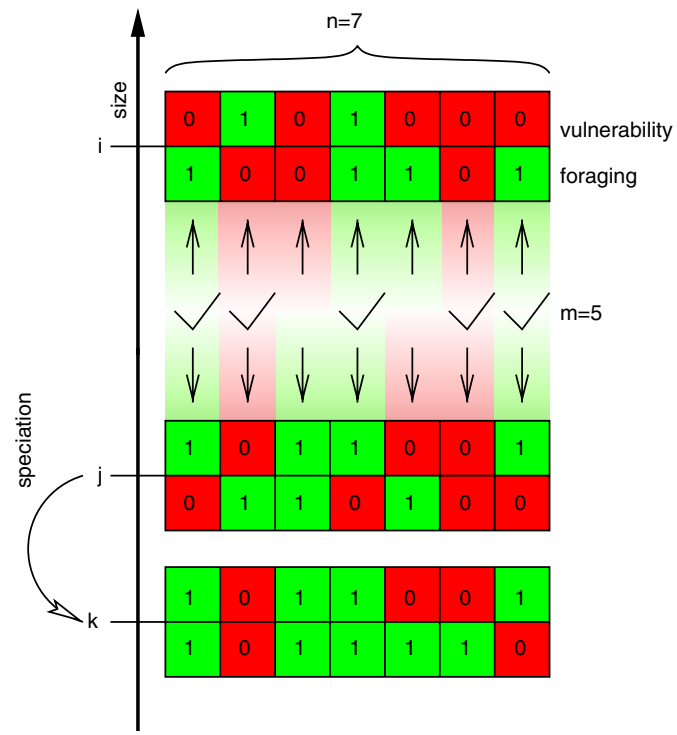


Fig. 1. The main components of the matching model. Each species (i, j, k) is characterized by n foraging and n vulnerability traits and a size parameter. Typically consumers (i) are larger than their resources (j). If the number m of matches between a consumer's foraging traits and a resource's vulnerabilities is large, trophic links result. In speciations ($j \to k$) some traits mutate. Foraging traits typically mutate more frequently than vulnerability traits. See text for details.

model parameter λ ($0 \leqslant \lambda \leqslant 1$) controls the amount of trophic loops (Polis, 1991) in a food web.

The complex processes driving evolution are modeled by speciations and extinctions that occur for each species randomly at rates r_+ and r_-, respectively (Raup, 1991). New species invade the habitat at a rate r_1. Such continuous-time birth-death processes are well understood (Bailey, 1964). With $r_+ < r_-$ the steady-state average of the number of species is $r_1/(r_- - r_+)$. For new, invading species the $2n$ traits and the size parameter s are determined at random with equal probabilities. For the descendant species of a speciation (Fig. 1), each vulnerability trait is flipped with probability p_v, each foraging trait is flipped with probability p_f, and a zero-mean Gaussian random number δ ($\mathrm{var}\,\delta = D$) is added to the size parameter s of the predecessor.[2] Such a random, undirected model of the evolution of trophic traits becomes plausible if one assumes the trophic niche space to be in a kind of "occupation equilibrium": there are no large voids in niche space to be filled and no niche-space regions of particularly strong predation pressure to avoid—hence evolution of traits into all directions is equally likely.

The model has the adjustable parameters r_+, r_-, r_1, λ, m_0, p_v, p_f, and D (see Table 1). An important derived

[1]We understand a habitat as being given by a set of environmental conditions rather than a location.

[2]$s = 0, 1$ are treated as reflecting boundaries (Rossberg et al., 2006).

quantity, which illustrates the role of m_0 in the model, is the probability for the link strength m to exceed the threshold m_0,

$$C_0 := P(m \geqslant m_0) = 2^{-n} \sum_{m=m_0}^{n} \binom{n}{m}. \quad (1)$$

For large n, food-web structure and dynamics become independent of n, provided m_0 is adjusted such as to keep C_0 constant (Appendix A). Throughout this work $n = 2^8 = 256$ is used, which is large enough to approximate the limit $n \to \infty$ and still computationally feasible. The assumption that all traits have equal weight in determining the link strength is not essential for the model dynamics. Traits could have different weight and be a mixture of qualitative and quantitative traits. The crucial point (Appendix A) is that the central limit theorem applies in determining the link strength m from a large number of traits. The complexity of most foraging interactions in nature, we believe, makes it plausible to assume that the number of relevant traits is large. Of course, this has to be tested by a comparison with empirical data.

Even though the speciation model (Rossberg et al., 2005) differs in many aspects form the matching model, it can be shown (Appendix B) that several analytic results derived for the former (Rossberg et al., 2006) apply under certain conditions also to the latter. A result important for simulations is, for example, that the model reaches a steady state after about $T = r_+^{-1} \ln[r_-/(r_- - r_+)]$ unit times. In order to ensure independent steady-state samples, each sample was obtained after initiating the model with zero species and letting it run for more than $5T$. Fig. 2 shows typical connection matrices of randomly sampled steady-state model webs in comparison with empirical data. More samples of connection matrices in comparison with data and the niche model, as well as a movie illustrating the model dynamics are available as supplementary material.

Table 1
Model parameters and their theoretical range

Model parameter	Range
Invasion rate r_1	$r_1 > 0$
Speciation rate r_+	$r_+ \geqslant 0$
Extinction rate r_-	$r_- > r_+$
Size dispersion constant D	$D \geqslant 0$
Loopiness λ	$0 \leqslant \lambda \leqslant 1$
Number of foraging/vulnerability traits n	$n \geqslant 1$
Threshold for link assignment m_0	$0 \leqslant m_0 \leqslant n$
Trait flipping probabilities p_v, p_f	$0 \leqslant p_v, p_f \leqslant 1/2$

3. Method of model validation

The model validation procedure we employed in this work is an extension of a method we used earlier (Rossberg et al., 2005). Two main statistics were computed: a χ^2 statistic directly characterizing the goodness of fit of the model to the empirical data and the Akaike Information Criterion (AIC) which admits a systematic comparison of the goodness of fit between different models, since it takes differences in model variabilities and in the number of fitting parameters into account. Both statistics were computed at maximum likelihood estimates of the parameters, precisely, the parameter values that maximize the (estimated) likelihood of obtaining, among those model samples that agree with the empirical data in the number of species, a model sample which agrees with a given empirical data sets in 13 quantitative food-web characteristics. The number of species and the 13 quantitative properties were computed after standardization of both model and empirical data. A detailed account of the procedure is given hereafter.

3.1. Empirical food-web data

The data base of empirical food-webs used for validating the model was provided by Dunne and Martinez. The

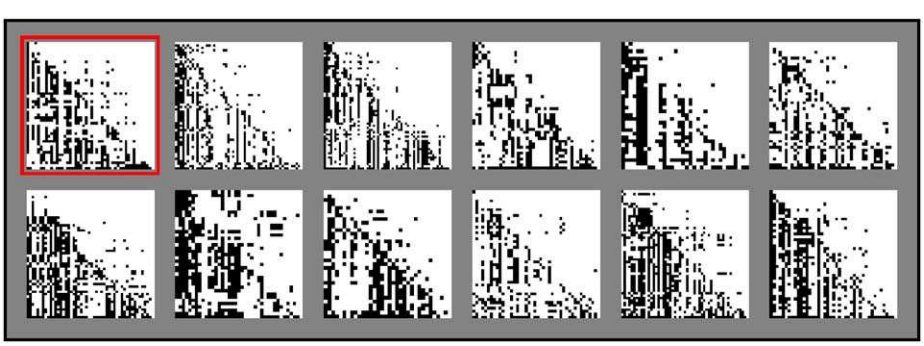


Fig. 2. Comparison between model steady state and empirical data. The connection matrix of the *Caribbean Reef* web (Opitz, 1996) (red box) is compared to the matrices of 11 random steady-state webs generated by the matching model (parameters as in Table 3). Each black pixel indicates that the species corresponding to its column eats the species corresponding to its row. Diagonal elements correspond to cannibalism. Pixel sizes vary due to varying webs sizes. For better comparison, data are displayed after standardization, a random permutation of all species, and a subsequent re-ordering such as to minimize entries in the upper triangle. Characteristic are, among others, the vertically stretched structures (Cattin et al., 2004) reflecting the strong inheritance of consumer sets.

following lists the labels we use for each data set and references to the original sources: *Benguela Current* (Yodzis, 1998), *Bridge Brook Lake* (Havens, 1992), *British Grassland* (Martinez et al., 1999), *Canton Creek* (Townsend et al., 1998), *Caribbean Reef* (Opitz, 1996), *Chesapeake Bay* (Baird and Ulanowicz, 1989), *Coachella Valley* (Polis, 1991), *El Verde Rainforest* (Waide and Reagen, 1996), *Little Rock Lake* (Martinez, 1991), *Northeast US Shelf* (Link, 2002), *Scotch Broom* (Memmott et al., 2000), *Skipwith Pond* (Warren, 1989), *St. Marks Seegrass* (Christian and Luczkovich, 1999), *St. Martin Island* (Goldwasser and Roughgarden, 1993), *Stony Stream* (Townsend et al., 1998), *Ythan Estuary 1* (Hall and Raffaelli, 1991), *Ythan Estuary 2* (Huxham et al., 1996).

3.2. Data standardization

Both empirical and model data were evaluated/compared after data standardization. The data standardization procedure consists of three steps:

1. Deleting disconnected species and small disconnected sub-webs. Graph theory predicts that there will be only a single large connected component. We keep only this large component.
2. Lumping of all species at the lowest trophic level into a single "trophic species". In the conventional procedure this step is omitted and only the lumping to "trophic species" (see next step) is performed. We added this step because in some data sets the lowest trophic level is particularly strongly lumped. For example, the *Chesapeake Bay* web contains a species "phytoplankton", and Coachella Valley "plants/plant products". On the other hand, food webs such as *Little Rock Lake* resolves the phytoplankton at the genus level. Lumping the lowest level improves data intercomparability.
3. The usual lumping of trophically equivalent species into single "trophic species" (Cohen et al., 1990).

For some data sets with a simple structure this procedure leads to a considerable reduction of the web size (e.g., *Bridge Brook Lake* shrinking from 74 species to 15). But generally this is not the case.

Obviously, the information that is lost in any of the three steps of data standardization does not enter the statistical analysis. This information is ignored because it is likely to be biased in at least some of the empirical data sets.

3.3. Food-web properties

Besides the number of species S, the following 13 food-web properties were used to characterize and compare empirical and model webs: the number of trophic links L expressed in terms of the directed connectance $C = L/S^2$ (Martinez, 1991), the clustering coefficient (Camacho et al., 2002; Dorogovtsev and Mendes, 2002) (*Clust* in Fig. 3); the fractions of cannibalistic species (Williams and Martinez, 2000) (*Cannib*) and species without consumers (Cohen et al., 1990) (T, top predators); the relative standard deviation in the number of resource species (Schoener, 1989) (*GenSD*, generality s.d.) and consumers (Schoener, 1989) (*VulSD*, vulnerability s.d.); the web average of the maximum of a species' Jaccard similarity (Jaccard, 1908) with any other species (Williams and Martinez, 2000) (*MxSim*); the fraction of triples of species with two or more resources, which have sets of resources that cannot be ordered to be all contiguous on a line (Cattin et al., 2004) (*Ddiet*); the average (Cohen et al., 1990) (*aChnLg*), standard deviation (Martinez, 1991) (*aChnSD*), and average per-species standard deviation (Goldwasser and Roughgarden, 1993) (*aOmniv*, omnivory) of the length of food chains, as well as the $\log_{10}$ of their total number (Martinez, 1991) (*aChnNo*). The prefix *a* at some property names indicates that these properties were computed using the fast, "deterministic" Berger–Shor approximation (Berger and Shor, 1990) of the maximum acyclic subgraph (MAS) of the food web, which makes these computations feasible also for large food webs. The number of non-cannibal trophic links not included in the MAS was measured as *aLoop*. When the output MAS of the Berger–Shor algorithm was not uniquely defined, the average over all possible outputs was used.

All food-web properties were calculated after data standardization as described above.

3.4. Fitting of the mean species number

Parameters where always chosen in such a way that the steady-state average $\overline{S}$ of the number of species S equals the value for a given empirical data set S_e (to within 8% for computational reasons). Practically this was achieved by adjusting a single model parameter (e.g. r_1 for the matching model or the number of species before lumping for the niche model), while keeping all other model parameters fixed. Since for the remaining parameters maximum-likelihood estimates were used and these are invariant under functional transformations, the specific choice of the parameter to be adjusted to archive $\overline{S} \approx S_e$ does not affect the final values of the best-fitting parameter set.

3.5. Sampling, averaging, and projection to $S = S_e$

The joint probability density of the 13 model food-web properties conditional to $S \approx S_e$ at fixed parameters was approximated by a multivariate normal distribution. To compute this distribution efficiently, we first obtained $N = 1000$ random samples of the model steady-state with S within 30% of S_e and computed the model averages and covariances of S and the other 13 food-web properties thereof. Then estimates of the model averages $\mathbf{v}$ and the covariance matrix $\mathbf{C}$ of the 13 food-web properties conditional to $S = S_e$ were obtained by a projection technique (essentially a linear regression, see Rossberg et al., 2005).

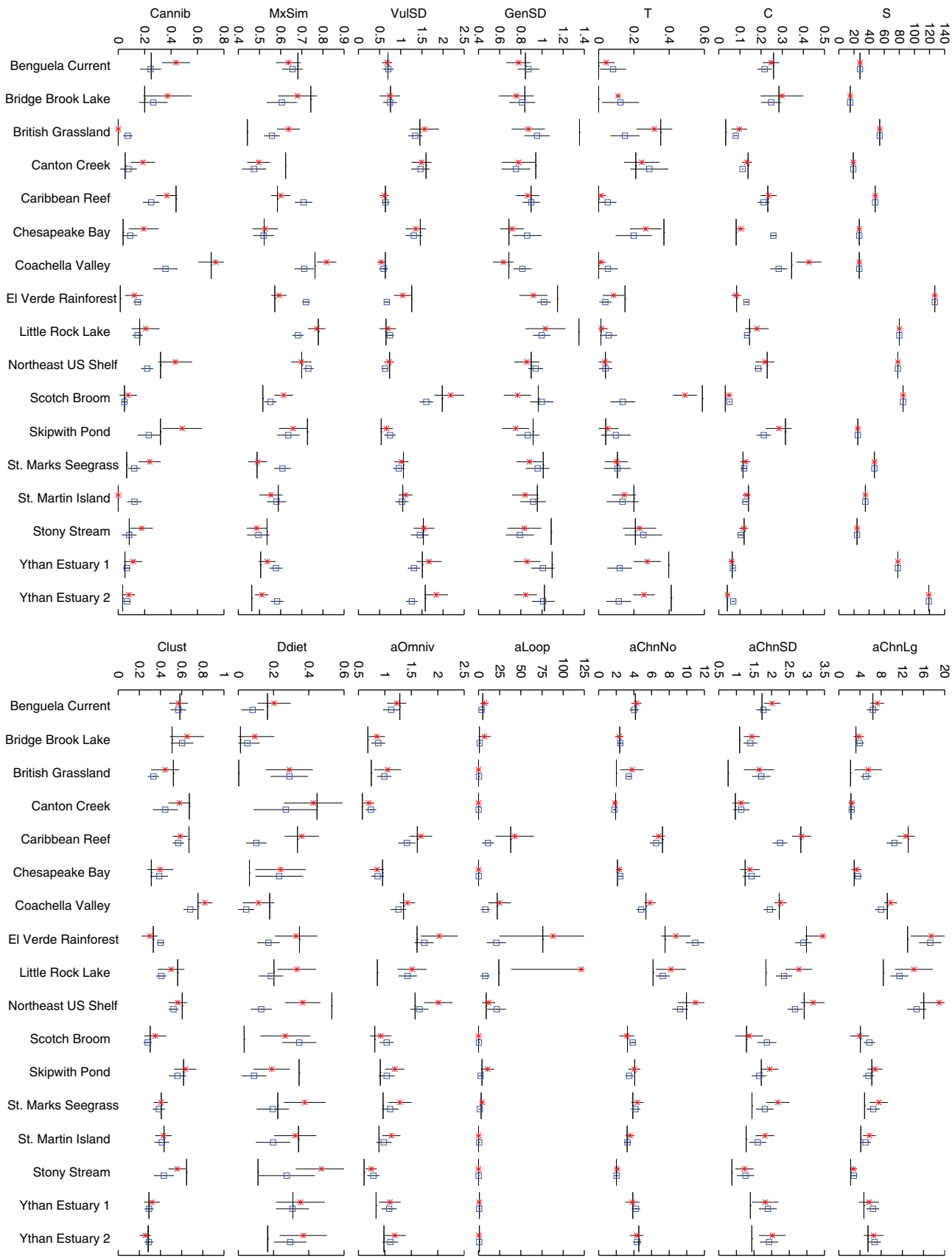


Fig. 3. Fitted food-web properties. The best fitting results for the matching model (red starts) and for the niche model (blue boxes) are compared to the empirical data (horizontal lines). Vertical lines correspond to ±1 model standard deviation. Because the properties are computed conditional to fixed S, the value of S always fits exactly.

The reason for computing the distribution at fixed S is that, for the small values of $1 - r_+/r_-$ that we find, the distribution of S broadens and deviates considerably from normality (Rossberg et al., 2006). Typically, less than a fifth of the sampled food webs had values of S within the 30% range around S_e.

3.6. Maximum likelihood estimation and AIC

Maximizing, for a given empirical data set, the likelihood that the vector $\mathbf{v}_e$ of empirical properties is reproduced by the model is equivalent to minimizing, in the normal approximation, the quantity

$$\Theta = \chi^2 + \ln|\mathbf{C}|, \tag{2}$$

with $\chi^2 := (\mathbf{v}_e - \bar{\mathbf{v}})^{\mathrm{T}} C^{-1} (\mathbf{v}_e - \bar{\mathbf{v}})$.

Since our numerical estimates for $\bar{\mathbf{v}}$ and $\mathbf{C}$ were based on a relatively small number N of steady-state samples, the numerical value of Θ itself had a sample standard deviation of about one.[3] The numerical minimization of Θ over the remaining model parameters was therefore carried out using the algorithm of Jones et al. (1998), which employs a Kriging technique to average and interpolation between numerical samples of Θ at different parameter values to obtain more accurate estimates of Θ and find its global minimum iteratively. The parameter values where the minimum was obtained are maximum likelihood estimates of the parameters for the given empirical data set.

We estimate the accuracy of the value $\Theta_{\min}$ we obtain for the global minimum of Θ, as computed by a Kriging interpolation between typically 200 numerical values (i.e. based on 2×10^5 Monte-Carlo samples) to be of the order 0.2. Finally, the AIC is obtained from $\Theta_{\min}$ as $\mathrm{AIC} = \Theta_{\min} + 2\times$ (number of adjustable parameters) + (some constant).

The same Kriging technique was also used to obtain improved estimates of χ^2, the 13 model properties, and their standard deviations at the likelihood maximum. Over all, a total of about 10^6 steady-state samples had to be computed for each empirical data set fitted.

4. Results of model validation

When applying this procedure to the matching model, only snapshots of the model steady state are compared with empirical data. Thus, only the relative evolution rates r_1/r_- and r_+/r_- matter. We set $r_- = 1$. The size-dispersion constant D has only a weak effect on results[4] and was kept fixed at $D = 0.05$. The remaining six parameters r_+, r_1, λ, m_0, p_v, and p_f were adjusted to fit 17 empirical data sets to the model as described above (Table 2). Model averages and standard deviations of the fitted food-web properties are compared with the empirical data in Fig. 3. Note, however, that this graphical comparison does not take covariances between food-web properties into account, and therefore gives only an incomplete account of the goodness of fit.

As a quantitative measure for the goodness-of-fit, the χ^2 values at the likelihood maximum (Section 3.6 above) were used. For a perfect fit under ideal conditions, this statistic has a χ^2-distribution with $14 - 6 = 8$ statistical degrees of freedom (DOF). The computed values are listed in Table 3 (χ^2_M).

Not all empirical food-webs are fitted equally well. For the three food webs labeled *Scotch Broom*, *British Grassland*, and *Ythan Estuary 2* the value of χ^2 exceeds the Bonferroni-corrected 95%-confidence limit $\chi^2 < 23.0$ (15 webs). Discrepancies between the remaining 14 data sets and the model, on the other hand, are revealed only when pooling all 14 sets: $\sum \chi^2 = 173$ for 112 DOF gives $p = 2 \times 10^{-4}$.

The niche model (Williams and Martinez, 2000) and the nested hierarchy model (Cattin et al., 2004), two of the best descriptions known so far, were fitted to the data using the same procedure.

As required for a systematic comparison between the models, the food webs generated by these model webs were standardized (Section 3.2), as well. Thus, model parameters had to be fitted numerically, deviating from the original prescriptions (Cattin et al., 2004; Williams and Martinez, 2000). Standardization of the model webs is suggested also for methodological consistency. For example, $Ddiet = 0$ for all raw niche-model food webs (Cattin et al., 2004), but can be non-zero after data standardization (Fig. 3). This might at first appear to be an undesired artifact of the standardization procedure. However, consider the hypothetical situation that a real food web is perfectly interval, but, due to the lumping of species during data recording, the published empirical network is not. Then, without applying our data standardization, *Ddiet*-values for the recorded empirical web and the niche-model would not match, even though in reality they do ($Ddiet = 0$). With data standardization this problem is, at least partially, remedied. Yet, data standardization retains information regarding the degree of intervality: As shown in Fig. 3, the value of *Ddiet* is lower for the niche model than it is for the matching model in most cases, even though the matching model itself has a tendency to generate interval food webs (Appendix B).

The χ^2 statistics for the niche model (χ^2_N) and the nested hierarchy model (χ^2_H) tend to be considerably higher than for the matching model (Table 3). Such a comparison does not take into account that these two models have only two parameters (corresponding to the number of species and the number of links) and thus four DOF more than the matching model. A simple statistic that does take the number of DOF into account is the number of "outliers" where χ^2 exceeds the Bonferroni-corrected 95% limit

[3]This numerical value agrees with an analytic estimate obtained from the difference of the variances of Hotelling's T^2 (divided by N) and the corresponding χ^2 distribution.

[4]For not too large λ, D, only the ratio $\lambda/D^{1/2}$ is relevant (Rossberg et al., 2006), as we verified numerically.

Table 2
Fitted parameters

Food-web name	r_1	$1-r_+/r_-$	λ	m_0	C_0[a]	p_v	p_f
Benguela Current	2.8	0.080	0.027	131	0.38	0.003	0.32
Bridge Brook Lake	1.4	0.033	0.13	136	0.17	0.000	0.068
British Grassland	1.4	0.014	0	139	0.09	0.014	0.013
Canton Creek	1.7	0.033	0.001	141	0.06	0.006	0.50
Caribbean Reef	0.48	0.0082	0.068	133	0.29	0.008	0.39
Chesapeake Bay	11.5	0.25	0.001	138	0.12	0.000	0.028
Coachella Valley	1.4	0.049	0.034	124	0.71	0.002	0.10
El Verde Rainforest	0.76	0.0054	0.12	139	0.09	0.015	0.036
Little Rock Lake	1.3	0.0092	0.25	138	0.12	0.001	0.043
Northeast US Shelf	0.28	0.0033	0.005	131	0.38	0.009	0.059
Scotch Broom	1.3	0.0067	0.001	144	0.03	0.031	0.006
Skipwith Pond	1.4	0.045	0.033	130	0.43	0.011	0.12
St. Marks Seegrass	0.55	0.0095	0.015	136	0.17	0.025	0.18
St. Martin Island	5.7	0.12	0	135	0.21	0.002	0.32
Stony Stream	0.13	0.0033	10^{-5}	141	0.06	0.014	0.35
Ythan Estuary 1	1.0	0.010	0.0029	140	0.08	0.033	0.037
Ythan Estuary 2	2.7	0.017	0.0022	141	0.06	0.041	0.038

[a]The linking probability C_0, is a derived quantity depending on m_0.

Table 3
Goodness of fit

Food-web name	χ^2_M	χ^2_N	$(\Delta\mathrm{AIC})_{M,N}$	χ^2_H	$(\Delta\mathrm{AIC})_{M,H}$
Benguela Current	14.1	14.1	11.6	48.3 *	−24.5
Bridge Brook Lake	12.0	12.4	15.2	30.9 *	−3.0
British Grassland	54.7 *[a]	144.6 *	−80.1	270.6 *	−181.2
Canton Creek	11.3	12.1	6.4	43.2 *	−24.3
Caribbean Reef	7.5	79.1 *	−52.8	168.8 *	−155.0
Chesapeake Bay	9.5	9.6	10.7	28.7	−8.5
Coachella Valley	5.9	31.9 *	−13.0	225.2 *	−207.1
El Verde Rainforest	14.0	337.6 *	−295.9	201.5 *	−165.7
Little Rock Lake	11.3	85.5 *	−46.8	227.5 *	−193.6
Northeast US Shelf	11.8	103.6 *	−73.9	204.5 *	−181.3
Scotch Broom	25.8 *	83.3 *	−42.5	309.3 *	−269.3
Skipwith Pond	14.3	39.9 *	−10.3	61.4 *	−31.8
St. Marks Seegrass	19.3	37.1 *	−3.5	22.3	8.2
St. Martin Island	7.4	13.7	−11.3	10.9	−11.0
Stony Stream	14.4	18.5	−3.5	45.9 *	−33.8
Ythan Estuary 1	20.5	42.0 *	−2.9	79.2 *	−43.6
Ythan Estuary 2	46.4 *	46.6 *	16.1	87.0 *	−29.5

[a]Stars (*) denote χ^2 values exceeding the Bonferroni-corrected 95% confidence limit.

(Table 3). For the matching model, there are three outliers as discussed above, for the niche model 11, and for the nested hierarchy model 14.

The low χ^2 obtained for the matching model are partially due to a larger variability of the model food webs. This effect and differences in the number of model parameters (the "model complexity") are both systematically taken into account by the AIC. However, only difference between AIC have a statistical meaning. The AIC was calculated for all three models (AIC_M, AIC_N, AIC_H). Negative values for $(\Delta\mathrm{AIC})_{M,N} := \mathrm{AIC}_M - \mathrm{AIC}_N$ and $(\Delta\mathrm{AIC})_{M,H} := \mathrm{AIC}_M - \mathrm{AIC}_H$ indicate that the matching model describes the data better than the niche model or the nested hierarchy model, respectively, and that the increased model complexity is justified. The matching model is favored over the niche model in 12 out of 17 cases, and over the nested hierarchy model in all but one case. In the cases where other models are favored, this is due to unnecessary complexity of the matching model, and not due to a better fit, as a comparison of the corresponding χ^2 values shows. Pooling all data yields $\sum(\Delta\mathrm{AIC})_{M,N} = -576$ and $\sum(\Delta\mathrm{AIC})_{M,H} = -1555$ in favor of the matching model.

We note that, when omitting the lumping of the lowest trophic level for data standardization, some χ^2-values increase for the matching model (AIC cannot be compared), such as for *Canton Creek* from 11.3 to about 23 or for *Northeast US Shelf* from 11.8 to about 45, while others, e.g. for *Bridge Brook Lake* or *St. Marks Seegrass*, are nearly unaffected. Overall, the changes are of similar size as the differences in the goodness of fit between the matching model and the other two models. Lumping the lowest level would thus not have much affected the quality of the fits of the older models, but at the level of accuracy reached here, it is crucial.

5. Discussion

Among the fitted parameters of the matching model, some depend just as much on methodological choices at the time of recording the food web as on the actual ecology. In particular, the linking threshold m_0 (or the linking probability C_0) is determined by the empirical threshold for link assignment, and the invasion rate r_1—as a parameter determining the web size—depends on the delineation of the habitat and the species-sampling effort. The degree of loopiness λ might depend on the particular method used to determine links empirically. Adjusting

these three parameters makes the model robust to differences in empirical methodology.

The remaining three parameters r_+, p_v, and p_f allow, at least partially, an ecological interpretation. r_+/r_- represents the fraction of species that entered the species pool by speciations from other species in the pool, in contrast to the remaining $1-r_+/r_-$ that entered through random "invasions". The low values found for $1-r_+/r_-$ (Table 2) indicate that evolutionary processes are essential for generating the observed structures.

The two quantities p_v and p_f measure the variabilities of vulnerability and foraging traits among related species. We typically find p_v much smaller than p_f (Table 3). In particular, $p_v < p_f$ in 14 of 17 data sets ($p = 0.006$). This implies that descendant species tend to acquire resources sets different from their ancestors but mostly share their enemies. We interpret this as a preference for avoiding resource competition rather than apparent competition (Holt and Lawton, 1994): a typical consumer is an expert for its particular set of resources (resource partitioning), and a typically resource set consists of a few "families" of related species—autotrophs or, again, expert consumers.

The three exceptional data sets with $p_v/p_f > 1$ are exactly those most difficult to fit by the matching model (Table 3). Interestingly, these are also the three data sets that contain large fractions (>30%) of parasites, parasitoids, and pathogens (PPP) in the resolved species pool. The other data sets are dominated by predators, grazers, and primary produces (PPP fraction $\lesssim 5\%$). These observations are consistent with the expectations that (i) due to their high specialization PPP are less susceptible to resource competition than predators (Morris et al., 2001) and (ii) the matching model does not describe PPP well because it assumes a size ordering which is typical only for predator–prey interactions (Cohen et al., 1993b; Memmott et al., 2000; Warren, 1989; Warren and Lawton, 1987). However, further investigations of these points are required. For example, contrary to expectations, p_v/p_f is close to one also for *Ythan Estuary 1*.

Another noteworthy observation is that all cases with positive $(\Delta\mathrm{AIC})_{M,N}$ correspond to aquatic systems. This might reflect the fact that the specificity of foraging, which our model describes in more detail than the niche model, is generally less pronounced in aquatic than in terrestrial system.

The matching model reproduces the empirical distributions of the numbers of consumers and resources of species well (see Fig. 4 and supplementary materials). Under specific conditions, which include $p_v \ll p_f$ (see Appendix A), these become the "universal", scaling distributions (Camacho et al., 2002; Stouffer et al., 2005) characteristic for the niche model (e.g., Fig. 4, *Caribbean Reef*). But the distributions for food webs deviating from these patterns are also reproduced (e.g., Fig. 4, *Scotch Broom*). The speciation model (Rossberg et al., 2005) could achieve this only under unrealistic assumptions regarding the allometric scaling of evolution rates (Rossberg et al., 2006).

As has been argued elsewhere (Cattin et al., 2004; Rossberg et al., 2006), the phylogenetic mechanism can also explain the long-known phenomenon of "intervality",

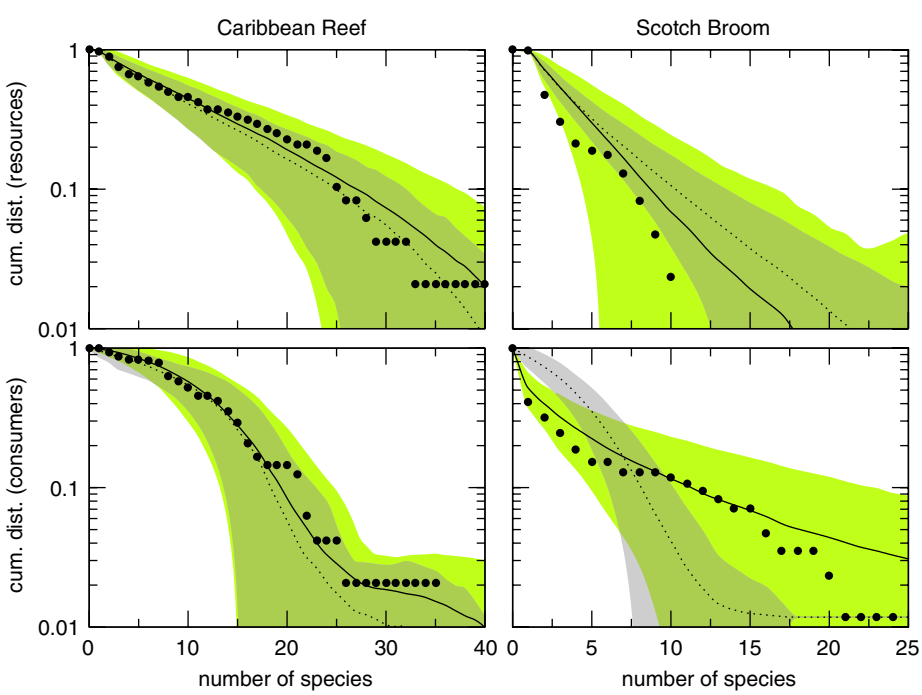


Fig. 4. Food-web degree distributions. Cumulative distributions for the number of resources (upper panels) and consumers (lower panels) of species for the *Caribbean Reef* and *Scotch Broom* webs after data standardization. Points denote empirical data, solid and dotted lines model averages for matching and niche model, respectively, obtained from those webs that agree with the empirical data in the number of species S. 2σ-ranges are indicated in green (matching model) and grey (niche model), olive at overlaps. Model parameters are as in Table 2.

that is, the observation that species can typically be ordered on a line in such a way that the diet of many consumer is a contiguous set (Cohen, 1978). According to the phylogenetic explanation, intervality reflects the fact that, in the ordering of species that is obtained by drawing them at the endpoints of the phylogenetic tree in the usual way, sets of closely related species are contiguous–independent of the precise definition of "closely related". Adding that consumers often feed on closely related species yields intervality (Rossberg et al., 2006).

The current analysis does not directly take information regarding the taxonomic or phylogenetic relationships between the species contained in the empirical data sets into account. We can therefore give only indirect evidence for the conclusion that the resource species of a consumer are typically closely related. But Cattin et al. (2004) have made a very similar observation based directly on taxonomic and food-web data: there is a close correlation between the degree of taxonomic similarity and the degree of trophic similarity. The results of Cattin et al. (2004) differ from ours in that their analysis symmetrically included links to consumers and links to resources for the definition of trophic similarity, while our analysis indicates that the trophic similarity is usually stronger for links to consumers (related resources have the same consumers) than for links to resources. A repetition of the analysis of Cattin et al. (2004) with an explicit distinction between consumer- and resource links might therefore provide an additional, independent test of our interpretation.

6. Conclusions

The surprisingly good fit of the matching model to the data of predatory food webs suggests that the model contains the essential mechanisms required to reproduce, explain, and predict, quantitatively and accurately, the empirical data. This is substantially more than a model that, based on some plausible mechanisms, qualitatively reproduces some observed phenomena.

As is well known from statistical theory, the accuracy of model validation is limited by (1) the amount of data used for the validation, (2) the degree of inherent variability of such data, (3) the quality of the data, and (4) the aspects of model and data that are compared (here the 13 network properties). The complexity and the ecological details that are reasonably incorporated into a model are, in turn, limited by the achievable degree of accuracy. Model selection criteria, such as the AIC, have been developed exactly for the purpose of balancing these limitations in a systematic and consistent way.

The effects incorporated in the matching model are not only justified by plausibility, but also by the data. As our computations indicate (not shown), dropping the size ordering ($\lambda = 1$) or enforcing it strictly ($\lambda = 0$), disabling phylogeny ($r_+ = 0$), disabling heredity ($p_v, p_f = \frac{1}{2}$) or rigidifying it ($p_v, p_f = 0$) would all typically worsen the AIC by several decades. On the other hand, the low χ^2-statistics obtained indicate that it will not be easy to improve the model further, provided the comparison based on the 13 network properties did not miss important aspects of the data.

We recall that increasing the complexity of a model will not generally improve its predictive power (nor the AIC), as it bears the risk of over-fitting parameters. Many conceivable model refinements such as going over from the box-car log-bodysize distribution used here to a more realistic shape (Blackburn and Gaston, 1994), or incorporating directed body-size evolution (Cope, 1887) in the model, very likely belong to this category.

The good fit between the matching model and empirical data goes against the intuition that the structure of food-webs is tightly related to population dynamical phenomena, such as complexity-stability relationships, Darwinian fitness, trophic cascades, or top–down and bottom–up effects. The matching model takes none of this into account. Two perspectives might help resolving this paradox: (i) It is conceivable that, even though food-web topology affects population dynamics, population dynamics has only a weak effect on topology. Why this reverse effect is so weak is not clear. One reason might be that the mechanisms by which population dynamics affects topology (e.g. by leading to extinctions or admitting invasions) are so complex that the correlations they mediate between the topology and changes in the topology are weak. But this hypothesis needs to be tested empirically or in population dynamical models. The matching model provides a clear statement of what needs to be shown. (ii) Presumably, there are some topological food-web properties which are not well reproduced by the matching model. For example, it appears that, just as the niche model, the matching model overestimates omnivory (Fig. 3). Population dynamics might explain such deviations (McCann, 2000). However, in order to be able to separate population-dynamical effects from the structures obtained from an undirected phylogeny, a good phylogenetic null model such as the matching model should be used for comparison.

The future probably belongs to carefully designed and validated models that describe both, evolutionary and population dynamics (Caldarelli et al., 1998; Drossel et al., 2001; Yoshida, 2003). A promising avenue might also be to take link-strengths explicitly into account in the characterization and validation of such models (Bersier et al., 2002; Hirata, 1995; Quince et al., 2005; Ulanowicz, 1997).

Acknowledgements

The authors thank J.A. Dunne and N.D. Martinez for making their food-web database available, T. Yamada for providing computational resources, N. Rajendran for insightful comments and discussion, and to The 21st Century COE Program "Environmental Risk Management for Bio/Eco-Systems" of the Ministry of Education, Culture, Sports, Science and Technology of Japan for financial support.

Appendix A. Derivation of link dynamics for large n

Here we explain why the network dynamics of the matching model becomes independent of n for large n, if m_0 is properly adjusted as n increases. First, consider a single trophic link from a (potential) consumer to a (potential) resource. Denote the foraging traits of the former by f_i, the vulnerability traits of the latter by v_i, where $i = 1, \ldots, n$ and $f_i, v_i \in \{0, 1\}$.

A.1. Linking probability

Consider the steady-state distribution of the link strength m defined by

$$m = \sum_{i=1}^{n} \left\{ \begin{array}{ll} 1 & \text{if } f_i = v_i \\ 0 & \text{if } f_i \neq v_i \end{array} \right\}. \tag{A.1}$$

Since the f_i and v_i are equally, independently distributed, m follows a binomial distribution with mean $n/2$ and standard deviation $\sigma = n^{1/2}/2$. The distribution of $x := (m - n/2)/\sigma$ converges to a standard normal distribution for large n. The linking probability C_0 converges to a fixed value $(2\pi)^{-1/2} \int_{x_0}^{\infty} \exp(-t^2/2)\,\mathrm{d}t$ if m_0 is adjusted such that $(m_0 - n/2)/\sigma$ converges to a fixed value x_0.

A.2. Link-strength mutation as an integrated Ornstein–Uhlenbeck process

In the following we argue that the dynamics of x between speciations can be characterized as an integrated Ornstein–Uhlenbeck process if n is large. First, consider only a single link, as above. When the resource speciates, its vulnerability traits are inherited by the descendant species, but with probability p_v they flip from v_i to $1 - v_i$. If $p_v < \frac{1}{2}$ this single step can be divided into a series of K small steps, where a property v_i is flipped in each step with a small probability q and otherwise left unchanged. Taking the possibility that properties are flipped repeatedly into account, one finds that the K small steps are equivalent to the speciation step if

$$p_v = \tfrac{1}{2}[1 - (1 - 2q)^K] \tag{A.2}$$

or

$$q = -\frac{\log(1 - 2p_v)}{2K} + \mathcal{O}(K^{-2}). \tag{A.3}$$

For sufficiently large K one has $qn \ll 1$. Then, at most one trait is flipped in each step, and the change in $x = (m - n/2)/\sigma$ is of order $\sigma^{-1} \sim n^{-1/2}$. As n increases, it becomes arbitrarily small.

Denote the value of m after the kth step by m_k. At each step, if m_k is known, the probability distribution of m_{k+1} depends only on n and m_k. If $qn \ll 1$, for example, one has $m_{k+1} = m_k - 1$ with probability $m_k q$, $m_{k+1} = m_k + 1$ with probability $(n - m_k)q$, and otherwise $m_{k+1} = m_k$. Thus, the dynamics of m—and of x—from step to step are Markov processes.

These three properties of the step-by-step dynamics of x in the limit of large n and K

1. normal distribution in the steady state,
2. Markov property,
3. arbitrarily small changes from step to step,

identify the dynamics as an Ornstein–Uhlenbeck process (Gardiner, 1990)

$$\mathrm{d}x(\tau) = -\mu\, x(\tau)\,\mathrm{d}t + \eta\,\mathrm{d}W(\tau), \tag{A.4}$$

where $W(\tau)$ is a Wiener process and $\tau = k/K$. In particular, one finds

$$\mu = -\log(1 - 2p_v), \quad \eta = \sqrt{2\mu}. \tag{A.5}$$

The value of x for a link from a speciating resource to its consumer is given by the integral of Eq. (A.4) over a τ-interval of unit-length, starting with the value of x for the ancestor. This implies that the correlation of x between direct relatives is $(1 - 2p_v)$ and between relatives of lth degree $(1 - 2p_v)^l$. The corresponding results for a speciating consumer are obtained by replacing p_v in Eq. (A.5) by p_f.

For the inheritance of several links to unrelated (hence uncorrelated) consumers, Eq. (A.4) holds for each link, and the Wiener processes are uncorrelated. For links to unrelated resources correspondingly. For links to related species the Wiener processes are correlated. From invariance considerations regarding the temporal ordering of evolutionary events in local networks one finds that for relatives of lth degree this correlation is $(1 - 2p_f)^l$ for species-as-consumers and $(1 - 2p_v)^l$ for species-as-resources. The correlations between links to related species from a newly invading species also follow this pattern. This provides a full characterization of the link dynamics for large n independent of n.

Appendix B. Relations between the matching model and the speciation model

In order to make the analytic characterizations of the degree distributions and other food-web properties obtained for the speciation model (Rossberg et al., 2006) accessible for the matching model, we derive an approximate description of the matching-model link dynamics that refers directly to the inheritance of connectivity between species, i.e., of the information if a link is present or not, rather than the inheritance of traits determining links.

Mathematically, this corresponds to a Markov approximation for the dynamics of the connectivity in the following form: if resource B speciates to C, its connectivity information to a consumer A is lost with a probability β_v (independent of the previous history) and otherwise copied

from B to C. When the information is lost, a link from C to A is established at random with probability C_0.

The breaking probability β_v can be obtained by equating the probabilities that A eats C given that A eats C's ancestor B for the exact description (in terms of p_v and m_0) and the Markov approximation. This gives

$$\beta_v = \frac{1}{2^n (1 - C_0) C_0} \sum_{m_1=m_0}^{n} \sum_{k=0}^{n-m_1} \sum_{m_2=k}^{m_0-1} \times \frac{n! p_v^{2k+m_1-m_2} (1-p_v)^{n-2k-m_1+m_2}}{k!\,(k+m_1-m_2)!(m_2-k)!(n-m_1-k)!} \tag{B.1}$$

with C_0 defined by Eq. (1). The corresponding expression for β_f is obtained by replacing p_v in Eq. (B.1) by p_f.

Many results of Rossberg et al. (2006) relied on the unrealistic assumption that consumers evolve much slower than their resources. This assumption was used to argue for

1. fully developed correlations of connectivity from one consumer to related resources and
2. absence of correlations for connectivity from one resource to related consumers.

Effects 1 and 2 were then used to simplify calculations. In the matching model 1 and 2 can be obtained without assuming large differences in speciation rates: effect 1 is obtained because statistical correlations in connectivity to related resources in the matching model depend only on the correlations between the traits of the resources, and not on the evolutionary history of the consumer (see also Appendix A). The correlations are large if p_v is small and, as a result, β_v is small. Effect 2 is obtained when p_f is close to 0.5 (foraging traits are randomized in speciations), which implies that β_f is close to 1.

Results of Rossberg et al. (2006) that contribute to a better understanding of the matching model include the computation of the time to reach the steady state, the derivation of the conditions under which the degree distributions become those of the niche model, and the explanation why model webs, just as empirical data (Cohen et al., 1990), exhibit a larger-than-random degree of "intervality". The average number of resource "families" (or "clades") of a consumer in the matching model can also be estimated, and turns out to be small: the largest value (3.7) is obtained for the top predator of *Ythan Estuary 2*. For most other webs this number is below two.

Appendix C. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.jtbi.2005.12.021.

References

Amaral, L.A.N., Meyer, M., 1999. Environmental changes, coextinction, and patterns in the fossil record. Phys. Rev. Lett. 82 (3), 652–655.

Bailey, N.T.J., 1964. The Elements of Stochastic Processes. Wiley, New York, pp. 97–101 (Chapter 8).

Baird, D., Ulanowicz, R.E., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. Ecol. Monogr. 59, 329–364.

Berger, B., Shor, P.W., 1990. Approximation algorithms for the maximum acyclic subgraph problem. In: Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, Philadelphia, pp. 236–243.

Berlow, E.L., Neutel, A.-M., Cohen, J.E., de Ruiter, P.C., Ebenman, B., Emmerson, M., Fox, J.W., Jansen, V.A.A., Jones, J.I., Kokkoris, G.D., Logofet, D.O., McKane, A.J., Montoya, J.M., Petchey, O., 2004. Interaction strengths in food webs: issues and opportunities. J. Anim. Ecol. 73, 585–598.

Bersier, L.-F., Dixon, P., Sugihara, G., 1999. Scale-invariant or scale-dependent behavior of the link density property in food webs: a matter of sampling effort? Proc. Natl Acad. Sci. 153, 676–682.

Bersier, L.-F., Banašek-Richter, C., Cattin, M.-F., 2002. Quantitative descriptions of food-web matrices. Ecology 83 (9), 2394–2407.

Blackburn, T., Gaston, K., 1994. Animal body size distribution: patterns, mechanisms and implications. Trends Ecol. Evol. 9, 471–474.

Briand, F., Cohen, J.E., 1984. Community food webs have scale-invariant structure. Nature 307, 264–267.

Caldarelli, G., Higgs, P.G., McKane, A.J., 1998. Modelling coevolution in multispecies communities. J. Theor. Biol. 193, 345.

Camacho, J., Guimerá, R., Amaral, L.A.N., 2002. Robust patterns in food web structure. Phys. Rev. Lett. 88 (22), 228102.

Cattin, M.-F., Bersier, L.-F., Banasek-Richter, C., Baltensperger, R., Gabriel, J.-P., 2004. Phylogenetic constraints and adaptation explain food-web structure. Nature 427, 835–839.

Christian, R.R., Luczkovich, J.J., 1999. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. Ecol. Modelling 117, 99–124.

Cohen, J.E., 1978. Food Webs and Niche Space. Princeton University Press, Princeton, NJ.

Cohen, J.E., Briand, F., Newman, C.M., 1990. Community Food Webs: Data and Theory. Biomathematics, vol. 20, Springer, Berlin, and references therein.

Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong, K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O'Malley, R., Page, L.M., Patten, B.C., Pimm, S.L., Polis, G.A., Rejmánek, M., Schoener, T.W., Schoenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R.E., Warren, P.H., Wilbur, H.M., Yodzis, P., 1993a. Improving food webs. Ecology 74 (1), 252–258.

Cohen, J.E., Pimm, S.L., Yodzis, P., Saldana, J., 1993b. Body sizes of animal predators and animal prey in food webs. J. Anim. Ecol. 62, 67–78.

Cope, E.D., 1887. The Origin of the Fittest. Appleton & Co, New York.

Dorogovtsev, S.N., Mendes, J.F.F., 2002. Evolution of networks. Adv. Phys. 51 (4), 1079–1187.

Drossel, B., Higgs, P.G., McKane, A.J., 2001. The influence of predator–prey population dynamics on the long-term evolution of food web structure. J. Theor. Biol. 208, 91–107.

Gardiner, C.W., 1990. Handbook of Stochastic Methods, second ed. Springer, Berlin.

Garlaschelli, D., Caldarelli, G., Pietronero, L., 2003. Universal scaling relations in food webs. Nature 423, 165–168.

Goldwasser, L., Roughgarden, J., 1993. Construction and analysis of a large Caribbean food web. Ecology 74, 1216–1233.

Hall, S.J., Raffaelli, D., 1991. Food-web patterns: lessons from a species-rich web. J. Anim. Ecol. 60, 823–842.

Havens, K., 1992. Scale and structure in natural food webs. Science 257, 1107–1109.

Hirata, H., 1995. Information theory and ecological networks. In: Jorgensen, S.E., Auerbach, S.I. (Eds.), Complex Ecology. The Part-Whole Relation in Ecosystems. Prentice-Hall PTR, Englewood Cliffs, NJ, USA, pp. 623–642.

Holt, R.D., Lawton, J.H., 1994. The ecological consequences of shared natural enemies. Annu. Rev. Ecol. Syst. 25, 495–520.

Huxham, M., Beany, S., Raffaelli, D., 1996. Do parasites reduce the chances of triangulation in a real food web? Oikos 76, 284–300.

Ito, H.C., Ikegami, T., 2006. Food-web formation with recursive evolutionary branching. J. Theor. Biol. 238(1), 1–10.

Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaudoise Sci. Nat. 44, 223–270.

Jones, D.R., Schonlau, M.S., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. J. Global Optim. 13 (4), 455–492.

Link, J., 2002. Does food web theory work for marine ecosystems? Mar. Ecol. Prog. Ser. 230, 1–9.

Loeuille, N., Loreau, M., 2005. Evolutionary emergence of size-structured food webs. Proc. Natl Acad. Sci. 102 (16), 5761–5766.

Martinez, N.D., 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. Ecol. Monogr. 61, 367–392.

Martinez, N.D., Hawkins, B.A., Dawah, H.A., Feifarek, P.B., 1999. Effects of sampling effort on characterization of food-web structure. Ecology 80 (3), 1044–1055.

May, R., 1999. Unanswered questions in ecology. Phil. Trans. R. Soc. Lond. 354, 1951–1959.

McCann, K.S., 2000. The diversity-stability debate. Nature 405, 228–233.

Memmott, J., Martinez, N.D., Cohen, J.E., 2000. Predators, parasitoids and pathogens: species richness, trophic generality and body size in a natural food web. J. Anim. Ecol. 69, 1–15.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: simple building blocks of complex networks. Science 298, 824–827.

Morris, R.J., Müller, C.N., Godfray, H.C., 2001. Field experiments testing for apparent competition between primary parasitoids mediated by secondary parasitoids. J. Anim. Ecol. 70, 301–309.

Neutel, A.-M., Heesterbeek, J.A.P., deRuiter, P.C., 2002. Stability in real food webs: weak links in long loops. Science 296, 1120–1123.

Opitz, S., 1996. Trophic interactions in Caribbean coral reefs. ICLARM Technical Report 43 43, 341.

Polis, G.A., 1991. Complex desert food webs: an empirical critique of food web theory. Am. Nat. 138, 123–155.

Quince, C., Higgs, P.G., McKane, A.J., 2005. Topological structure and interaction strengths in model food webs. Ecol. Modeling, doi:10.1016/j.ecolmodel.2004.12.018.

Raup, D.M., 1991. A kill curve for Phanerozoic marine species. Paleobiology 17 (1), 37–48.

Rossberg, A.G., Matsuda, H., Amemiya, T., Itoh, K., 2005. An explanatory model for food-web structure and evolution. Ecol. Complexity 2, 312–321.

Rossberg, A.G., Matsuda, H., Amemiya, T., Itoh, K., 2006. Some properties of the speciation model for food-web structure—mechanisms for degree distributions and intervality. J. Theor. Biol. 238(2), 401–415.

Schoener, T.W., 1989. Food webs from the small to the large. Ecology 70, 1559–1589.

Stouffer, D.B., Camacho, J., Guimerà, R., Ng, C.A., Amarala, L.A.N., 2005. Quantitative patterns in the structure of model and empirical food webs. Ecology 86 (5), 1301–1311.

Thompson, R.M., Townsend, C.R., 2005. Food-web topology varies with spatial scale in a patchy environment. Ecology 86 (7), 1916–1925.

Tokita, K., Yasutomi, A., 2003. Emergence of a complex and stable network in a model ecosystem with extinction and mutation. Theor. Popul. Biol. 63, 131–146.

Townsend, C.R., Thompson, R.M., McIntosh, A.R., Kilroy, C., Edwards, E., Scarsbrook, M.R., 1998. Disturbance, resource supply, and food-web architecture in streams. Ecol. Lett. 1, 200–209.

Ulanowicz, R.E., 1997. Ecology, the Ascendent Perspective. Columbia University Press, New York.

Waide, R.B., Reagan, W.B. (Eds.), 1996. The Food Web of a Tropical Rainforest. University of Chicago Press, Chicago.

Warren, P.H., 1989. Spatial and temporal variation in the structure of a freshwater food web. Oikos 55, 299–311.

Warren, R.J., Lawton, J.H., 1987. Invertebrate predator–prey body size relationships: an explanation of upper triangularity in food webs and patterns in food web structure. Oecologia 74, 231–235.

Wilhelm, T., 2003. An elementary dynamic model for non-binary food webs. Ecol. Modeling 168, 145–152.

Williams, R.J., Martinez, N.D., 2000. Simple rules yield complex food webs. Nature 404, 180–183.

Yodzis, P., 1981. The stability of real ecosystems. Nature 289, 674–676.

Yodzis, P., 1998. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. J. Anim. Ecol. 67 (4), 635–658.

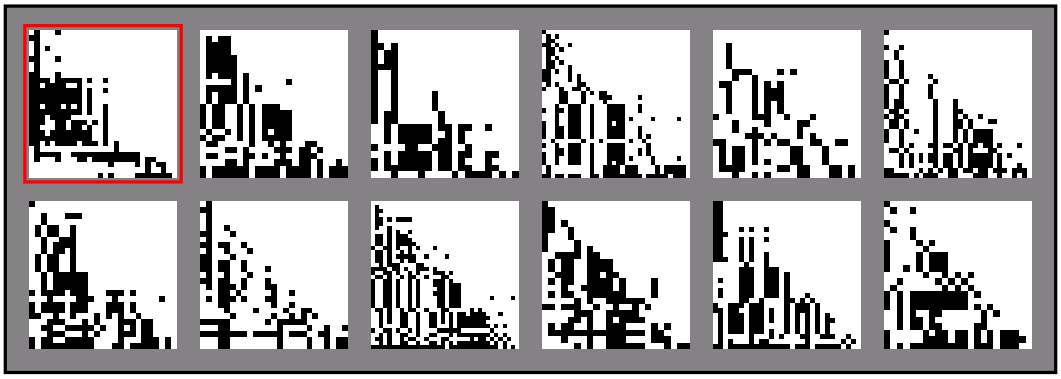
Yoshida, K., 2003. Dynamics of evolutionary patterns of clades in a food web system model. Ecol. Res. 18, 625–637.

More connection matrixes and degree distributions

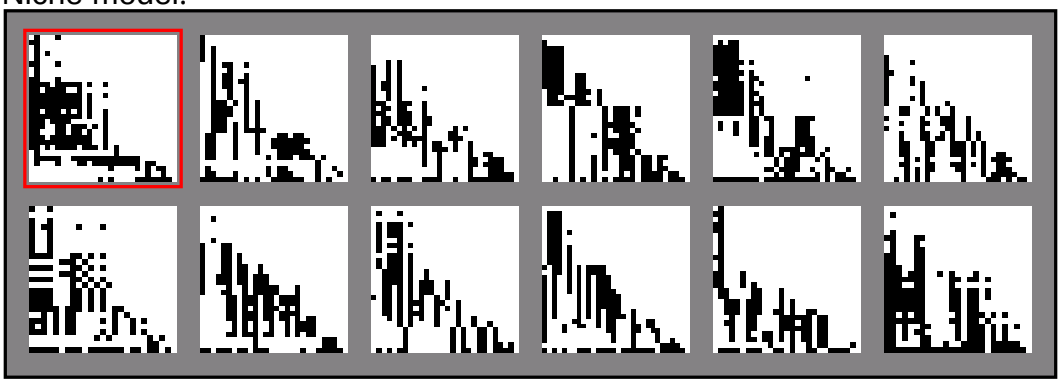
The figures in this file present the results corresponding to Figures 2 and 3 of the main text for the matching model (this work) and the niche model (Williams and Martinez 2000) for all empirical data sets considered. In each figure, the first panel shows the connection matrix of the empirical food web in a red box compared to the first 11 random samples obtained from a simulation of the matching model. Each black pixel indicates that the species corresponding to its column eats the species corresponding to its row. Diagonal elements correspond to cannibalism. Pixel sizes vary due to varying webs sizes. Data are displayed after standardization, a random permutation of all species, and a subsequent re-ordering such as to minimize entries in the upper triangle. The second panel in each figure displays the corresponding data for the niche model. The two bottom panels compare model and empirical degree distributions (model parameters as in Table 2). As in Figure 3 of the main text, points denote empirical data, and solid and dotted lines model averages for matching and niche model, respectively; 2σ-ranges are indicated in green (matching model) and grey (niche model), olive at overlaps. All model distributions were calculated conditional to S fixed at the empirical value.

Figure 1: Benguela Current

Matching model:



Niche model:



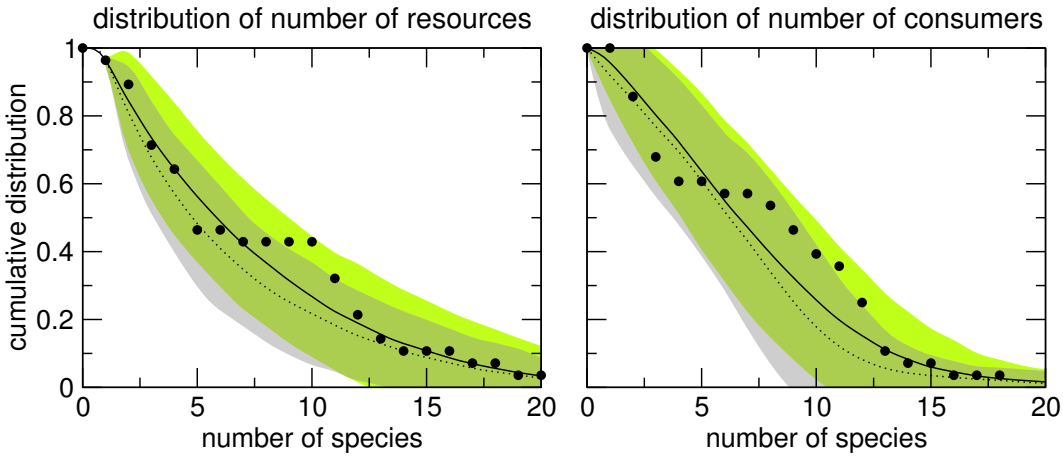
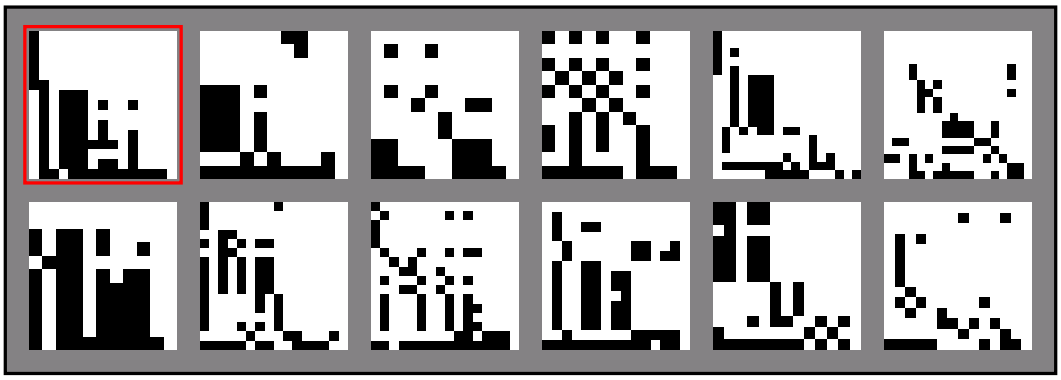
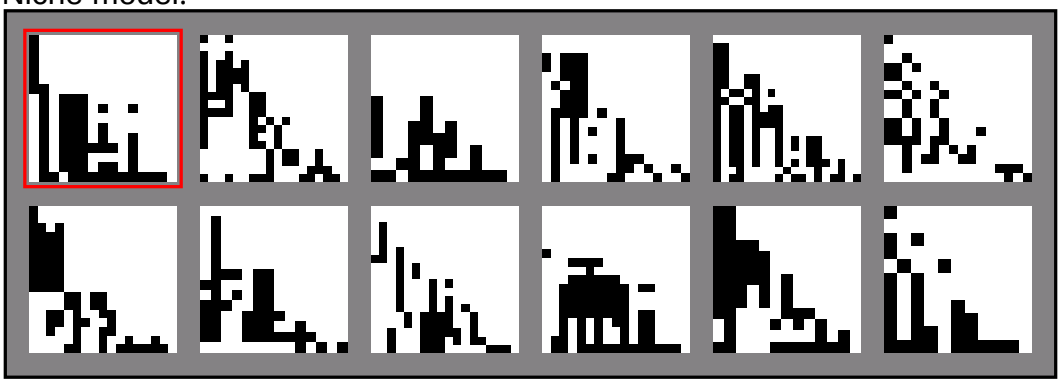


Figure 2: Bridge Brook Lake

Matching model:



Niche model:



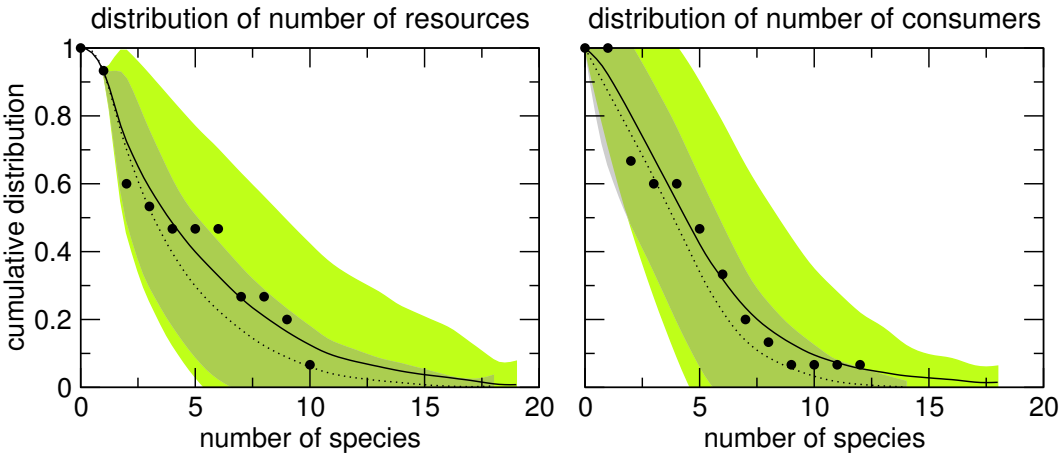
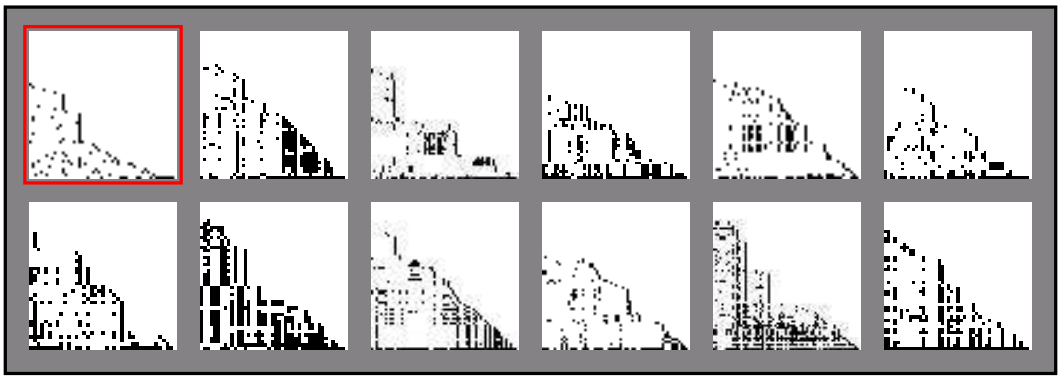
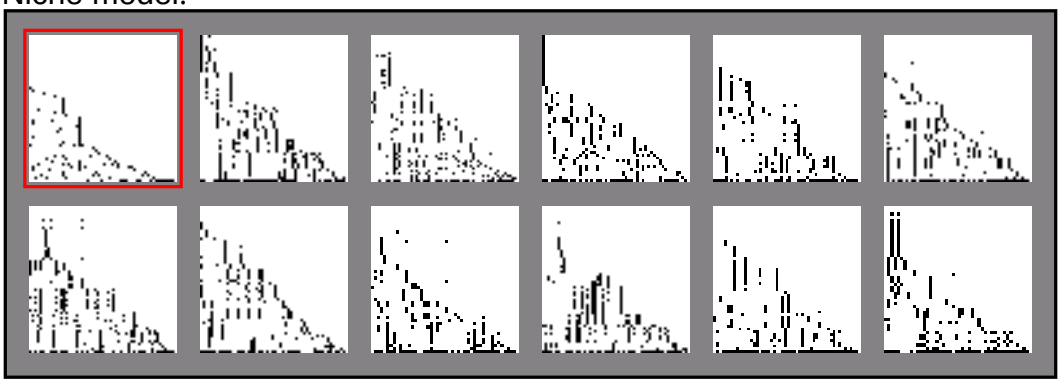


Figure 3: British Grassland

Matching model:



Niche model:



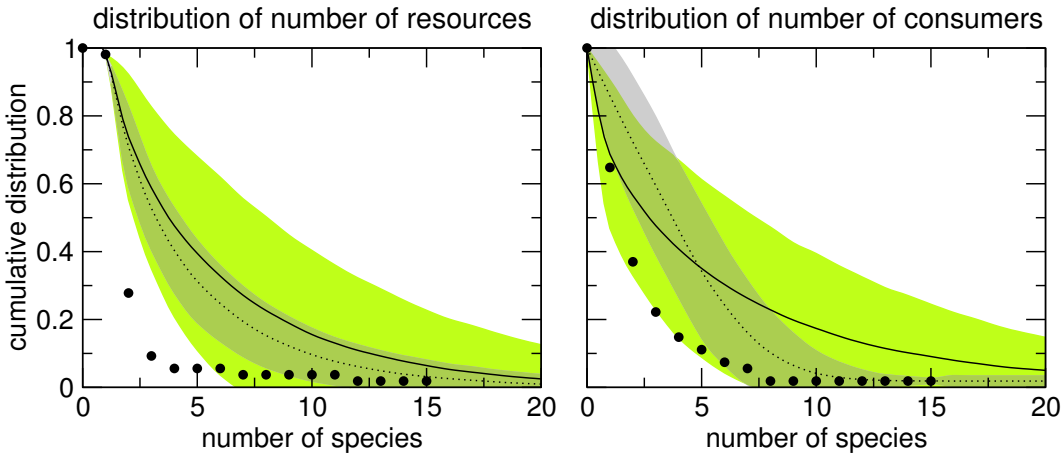
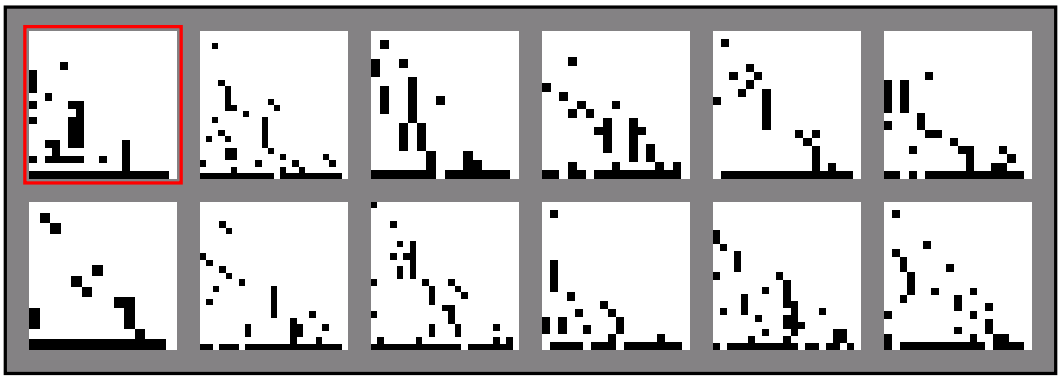
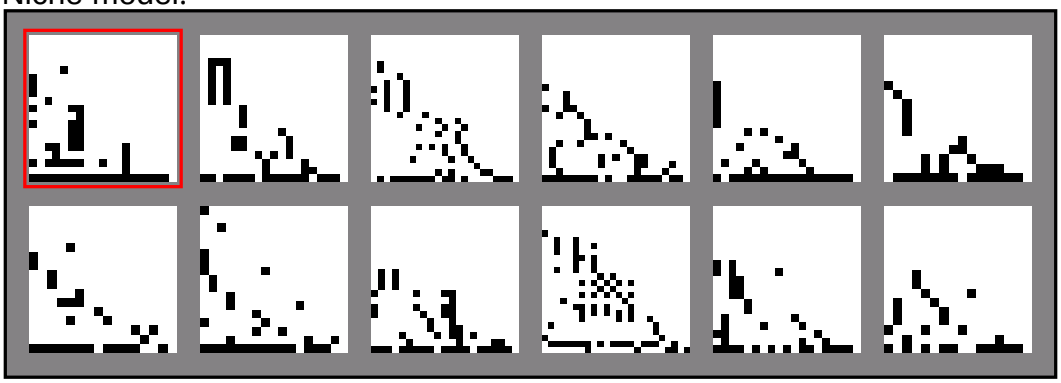


Figure 4: Canton Creek

Matching model:



Niche model:



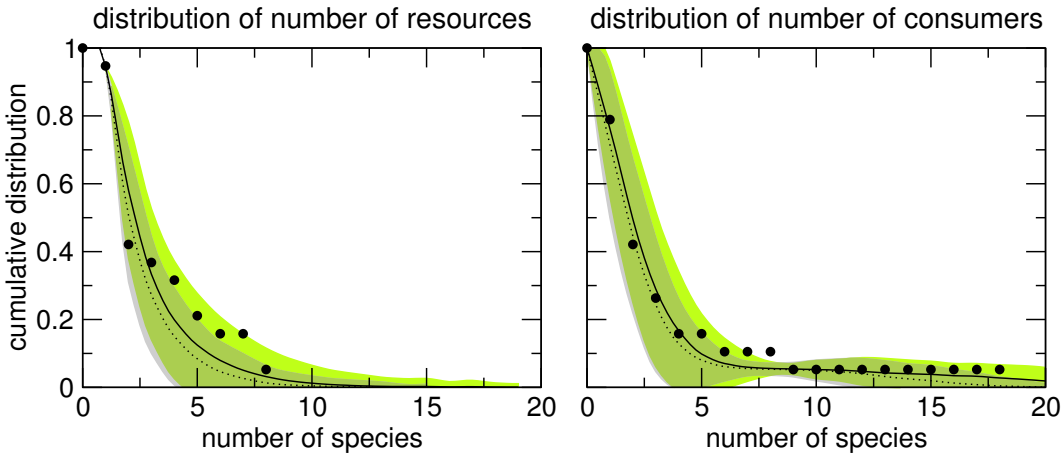
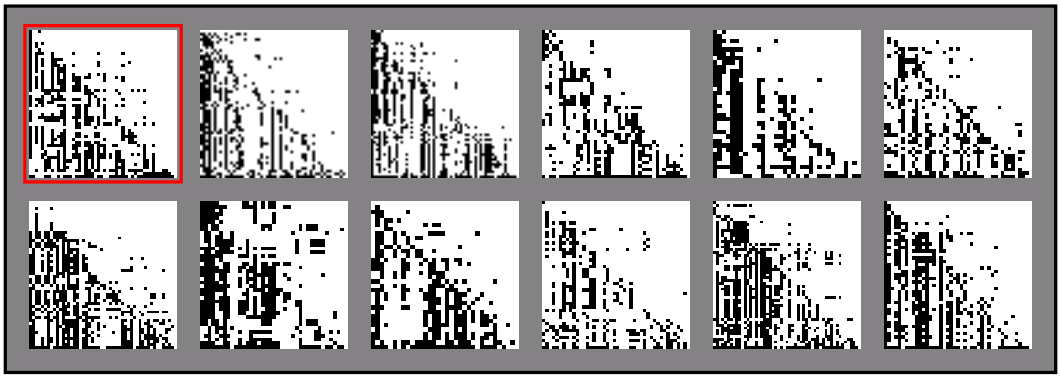
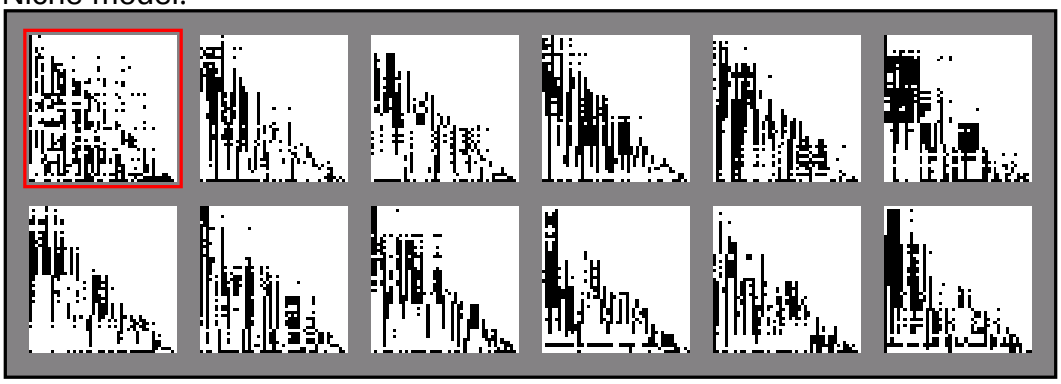


Figure 5: Caribbean Reef

Matching model:



Niche model:



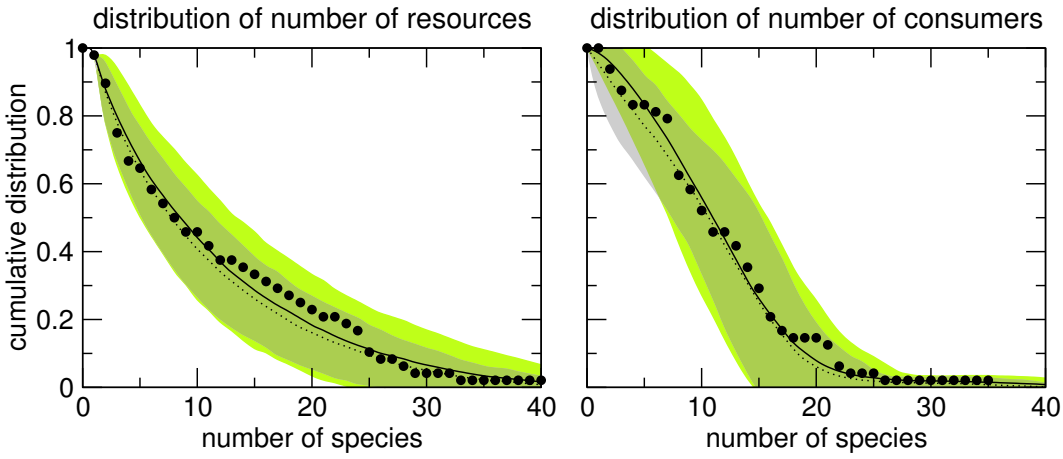
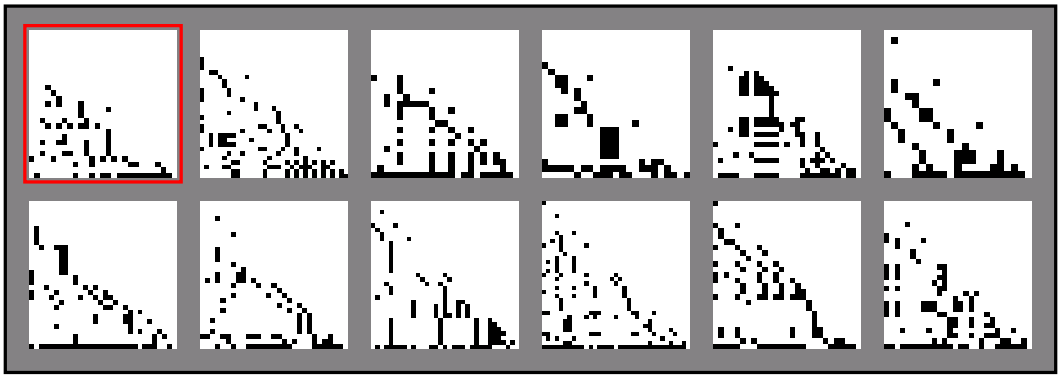
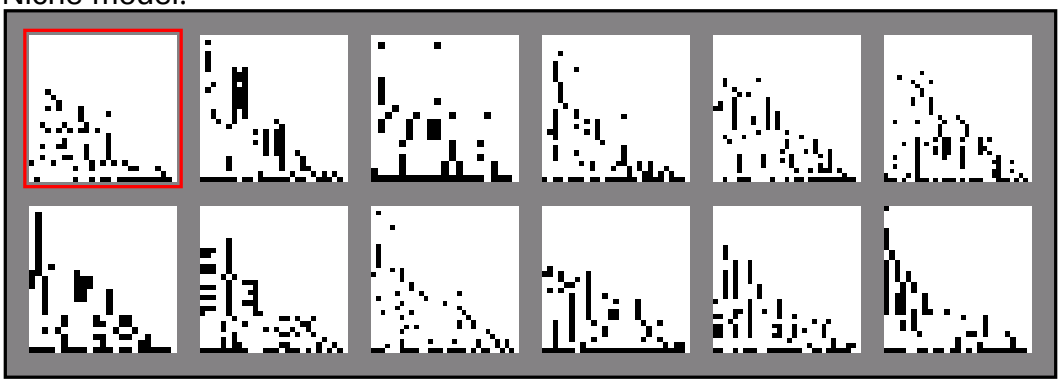


Figure 6: Chesapeake Bay

Matching model:



Niche model:



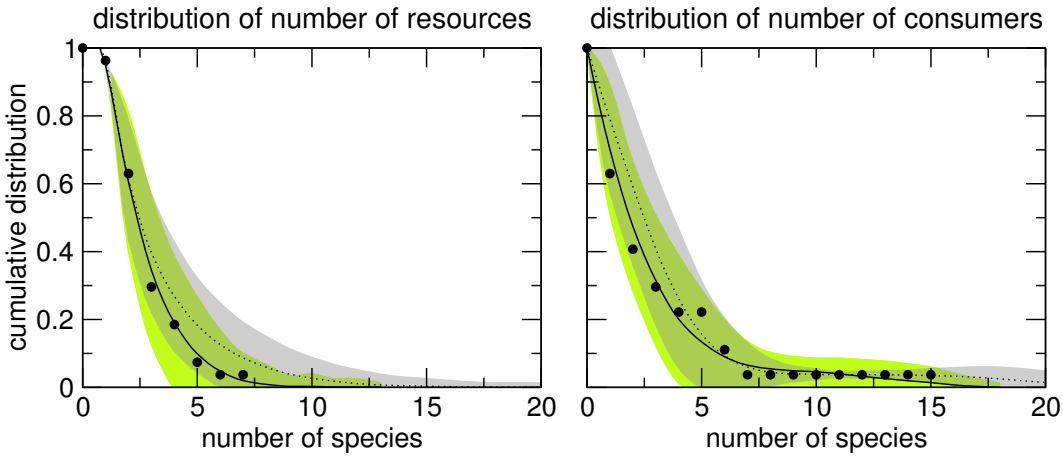
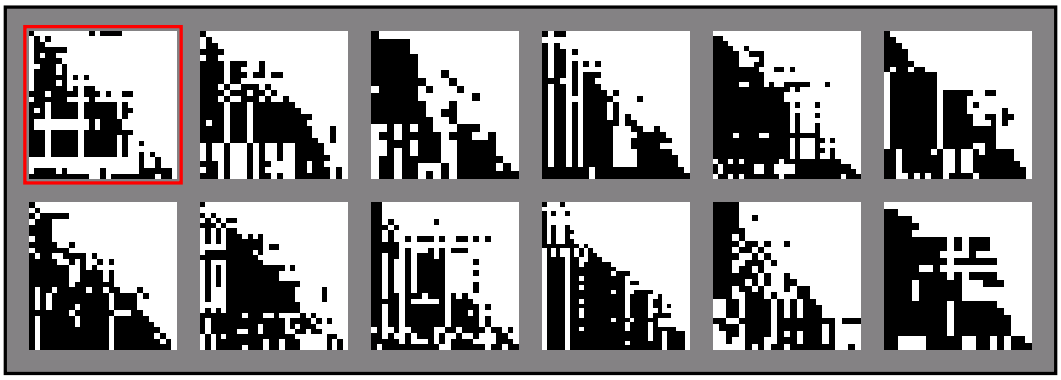
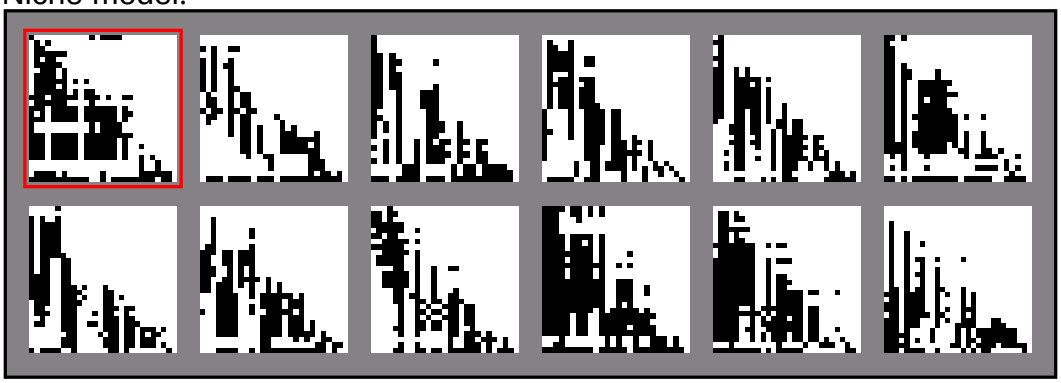


Figure 7: Coachella Valley

Matching model:



Niche model:



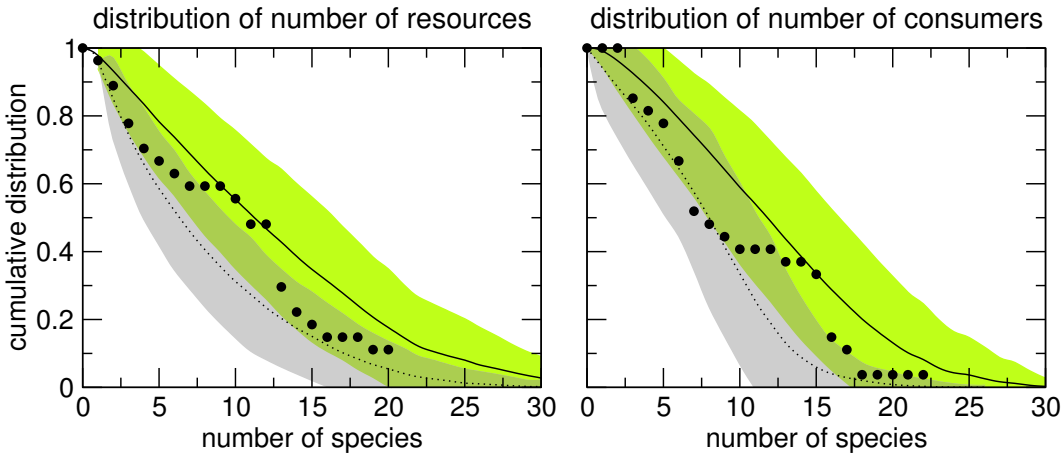
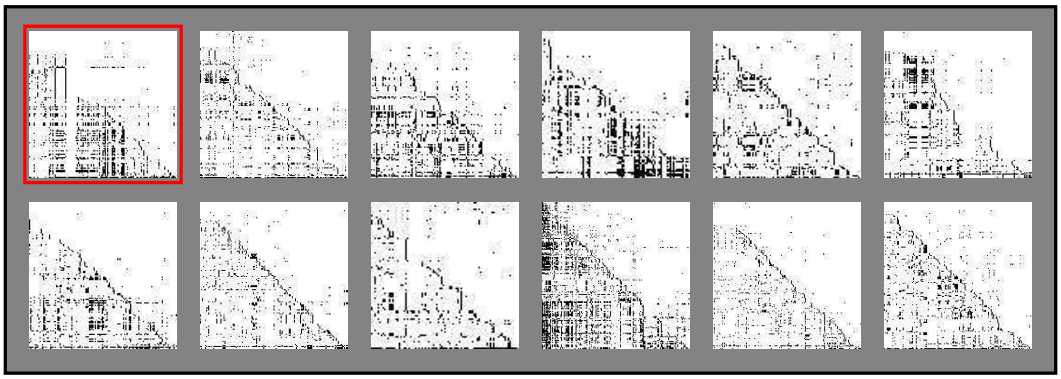
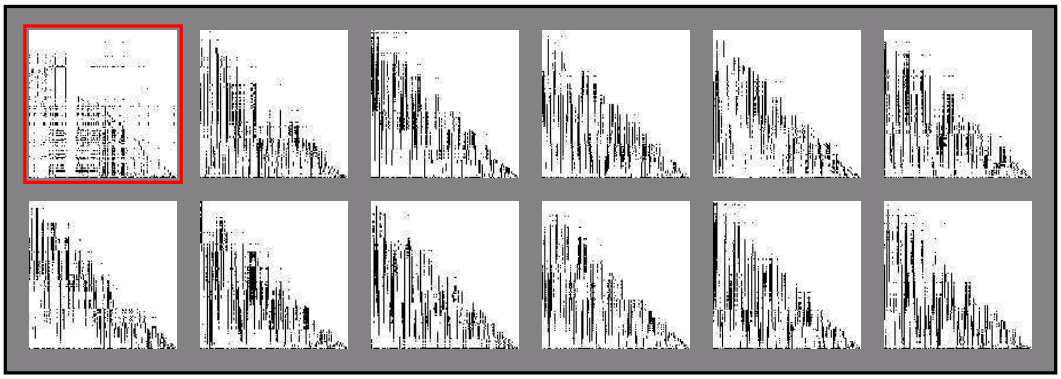


Figure 8: El Verde Rainforest

Matching model:



Niche model:



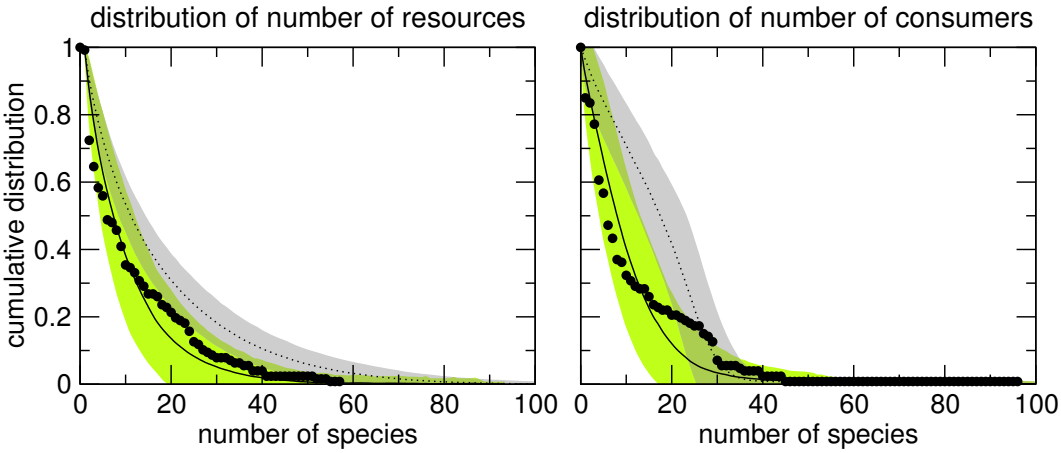
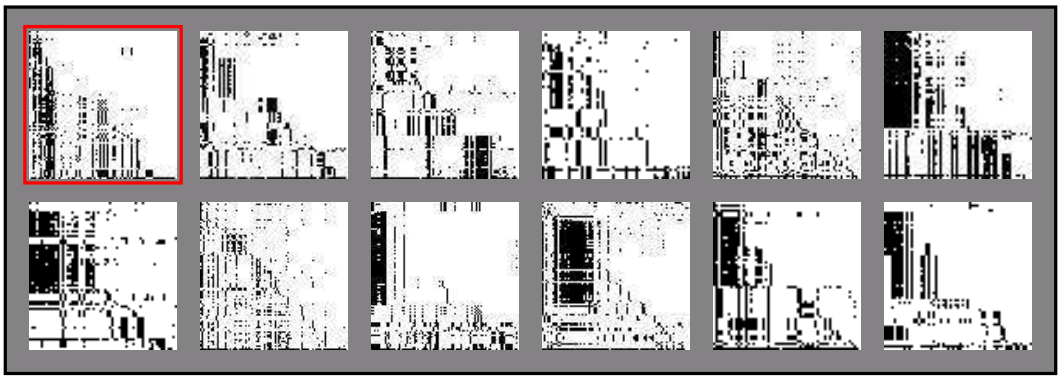
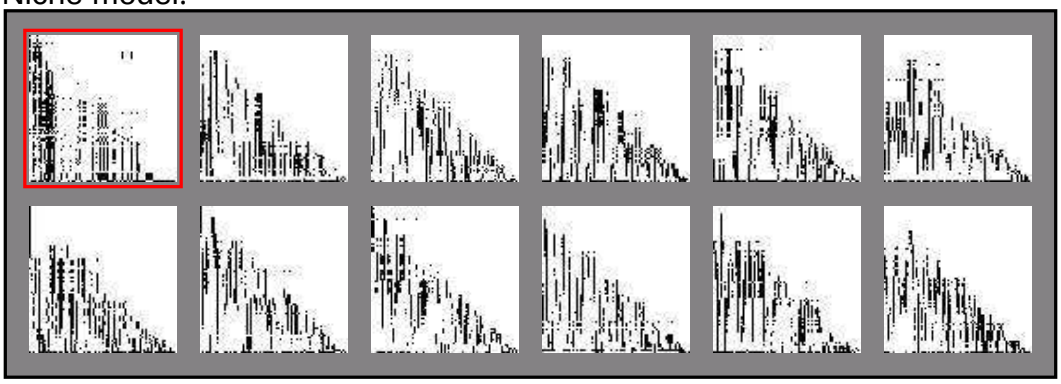


Figure 9: Little Rock Lake

Matching model:



Niche model:



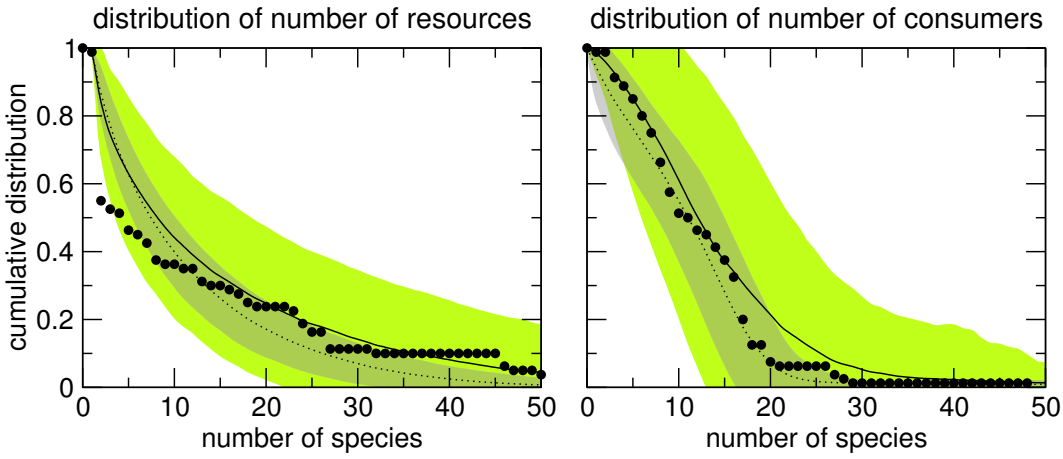
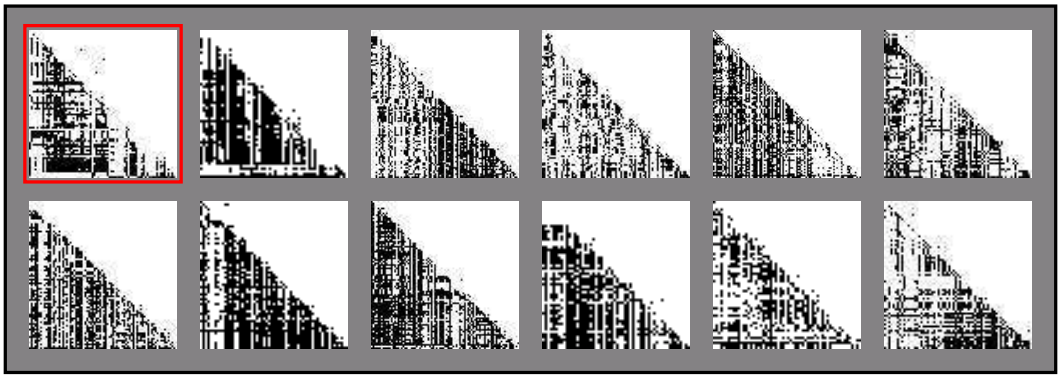
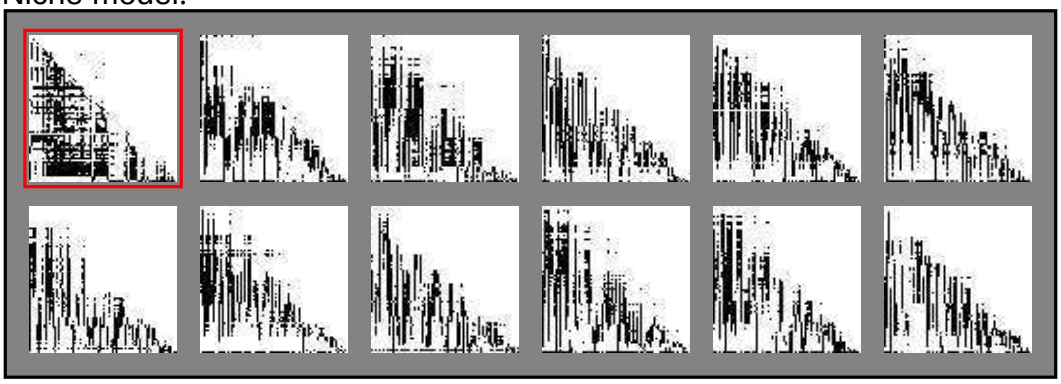


Figure 10: Northeast US Shelf

Matching model:



Niche model:



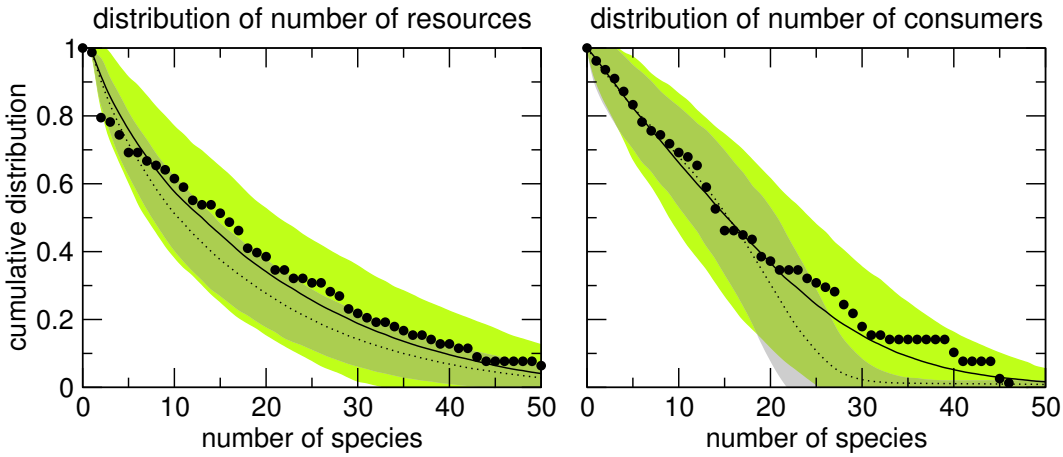
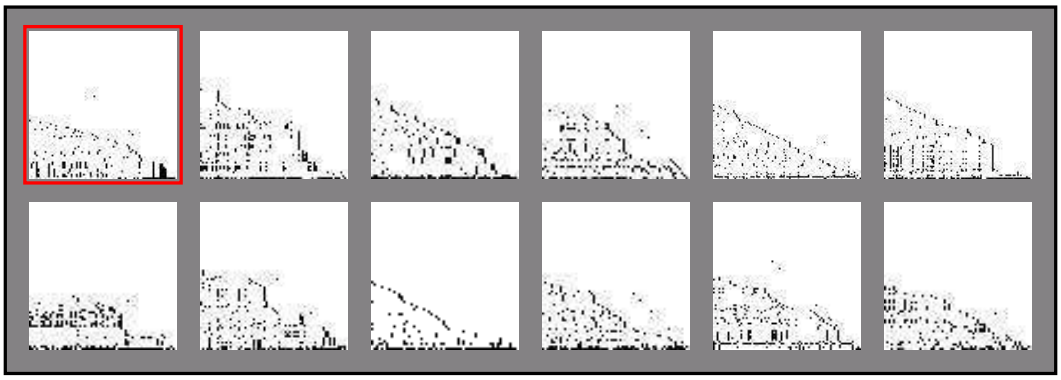
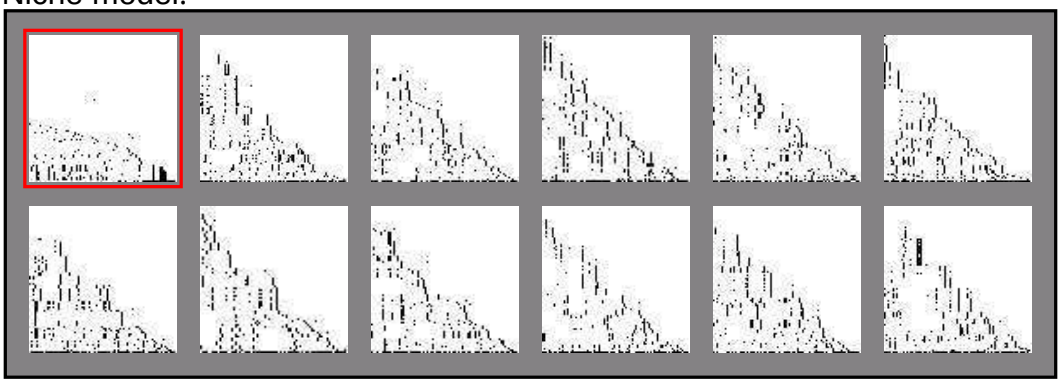


Figure 11: Scotch Broom

Matching model:



Niche model:



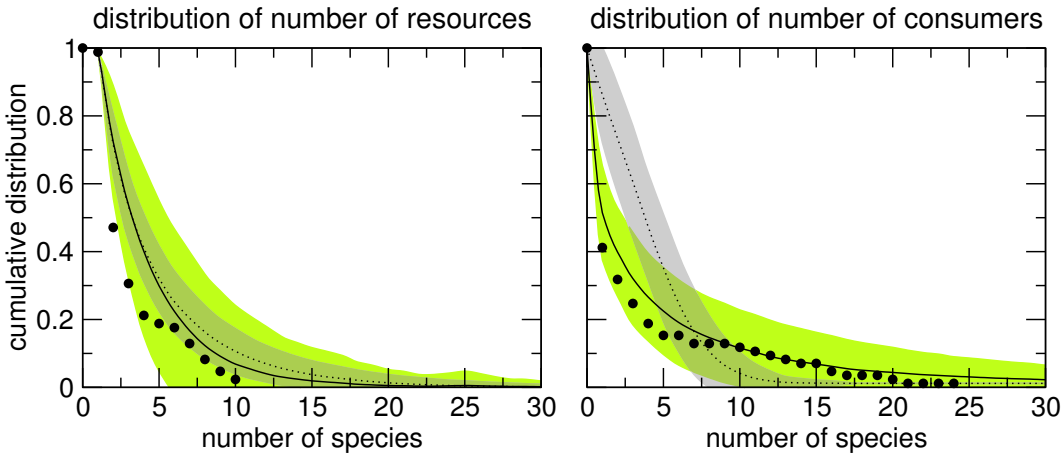
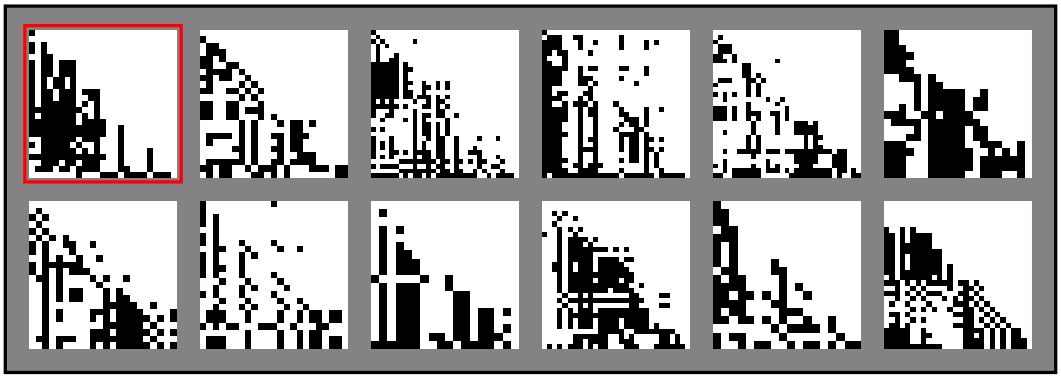
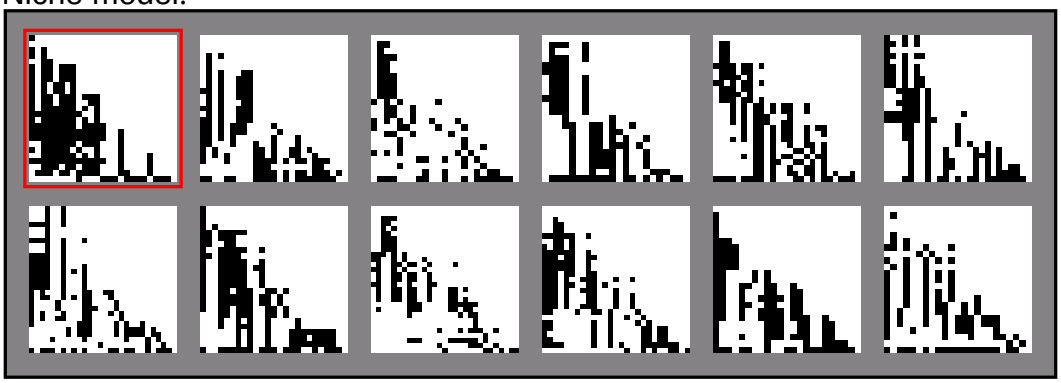


Figure 12: Skipwith Pond

Matching model:



Niche model:



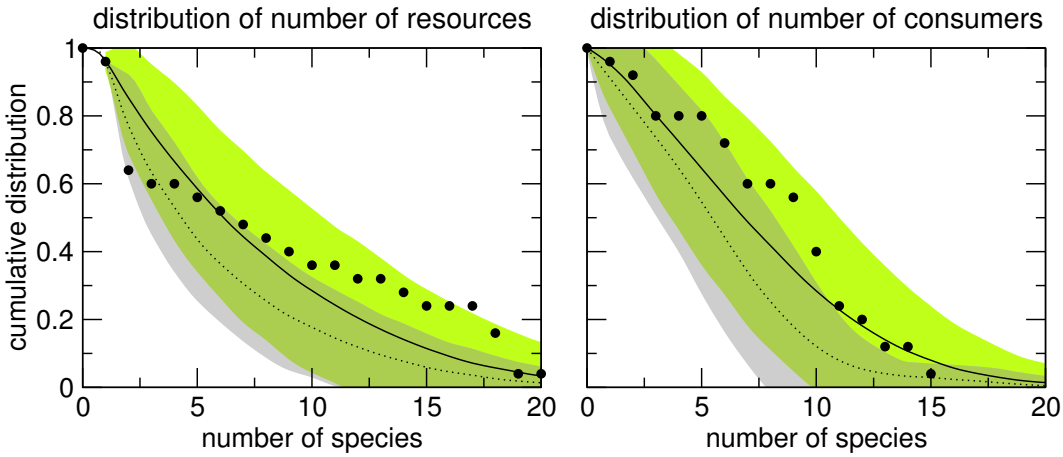
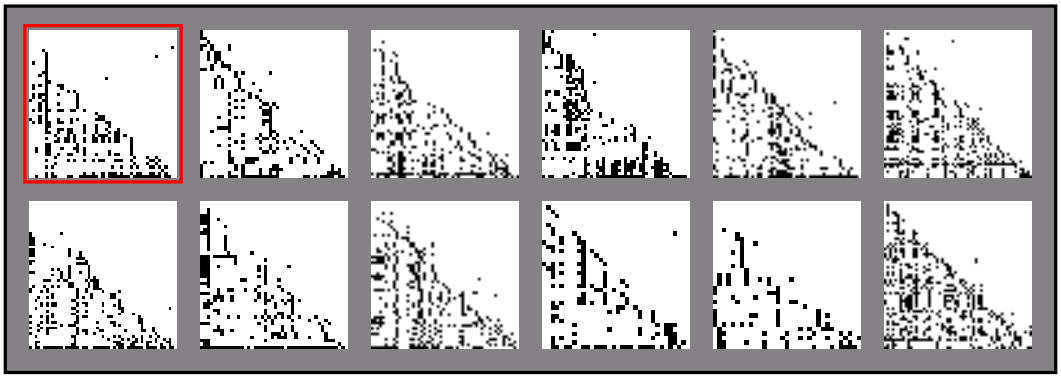
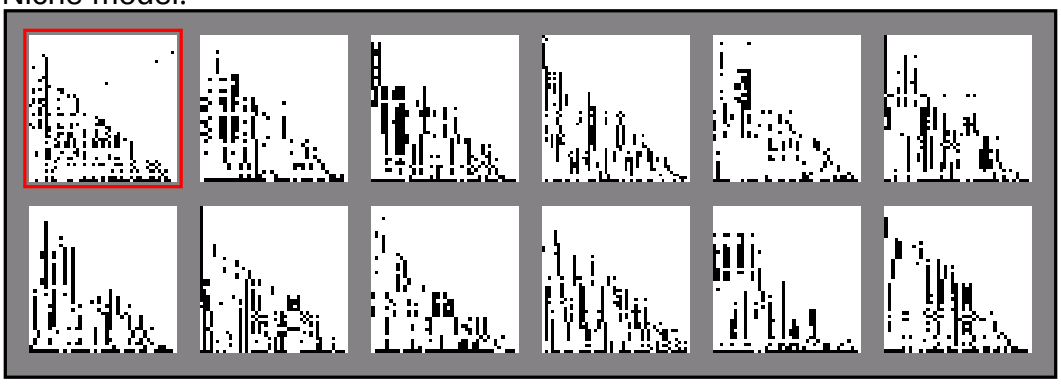


Figure 13: St. Marks Seegrass

Matching model:



Niche model:



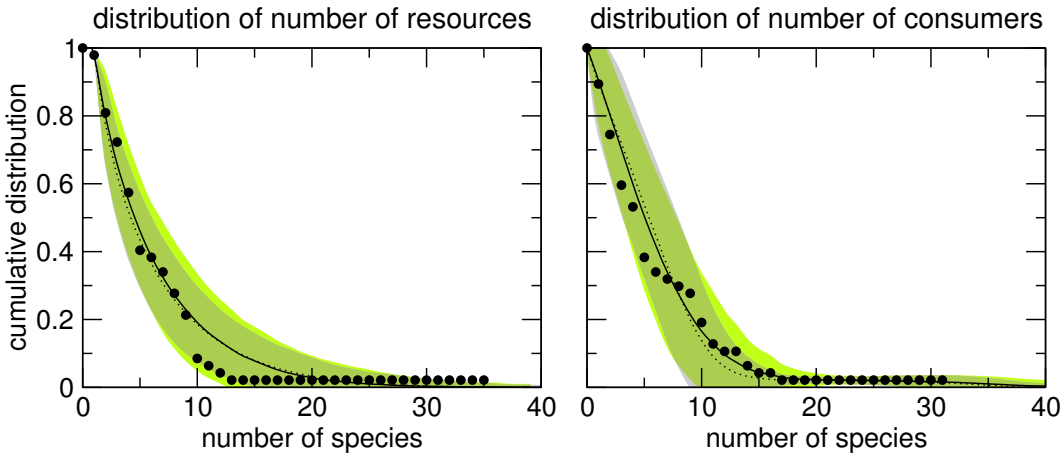
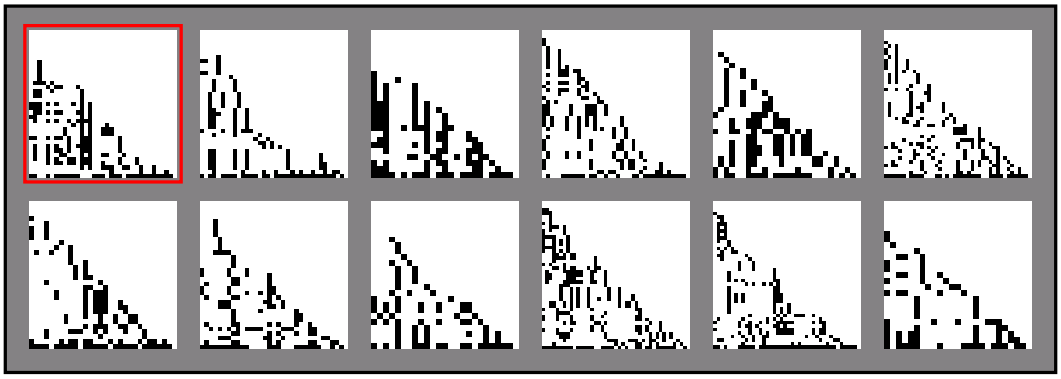
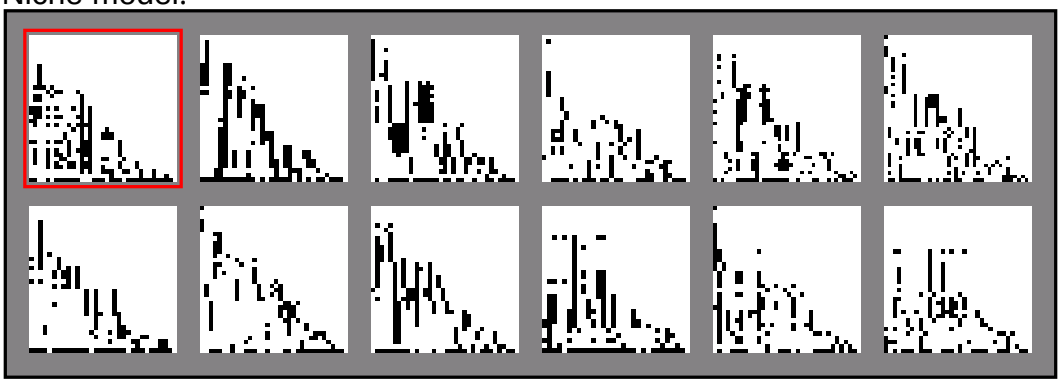


Figure 14: St. Martin Island

Matching model:



Niche model:



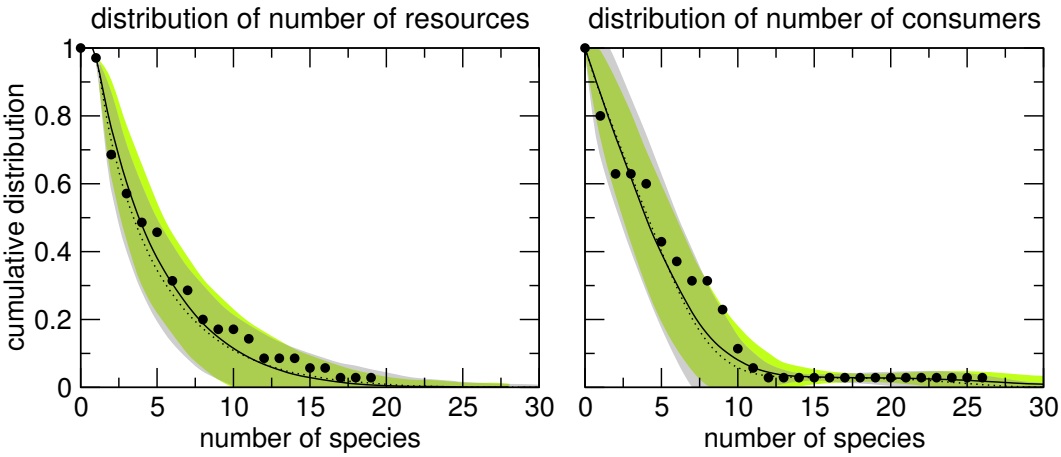
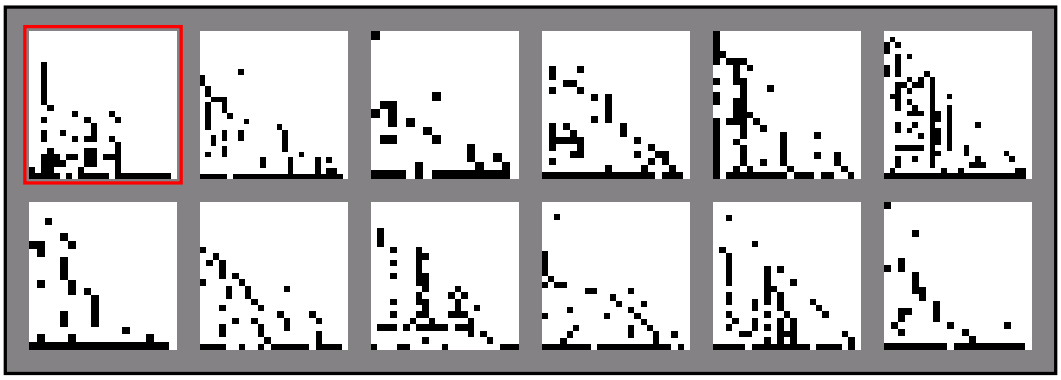
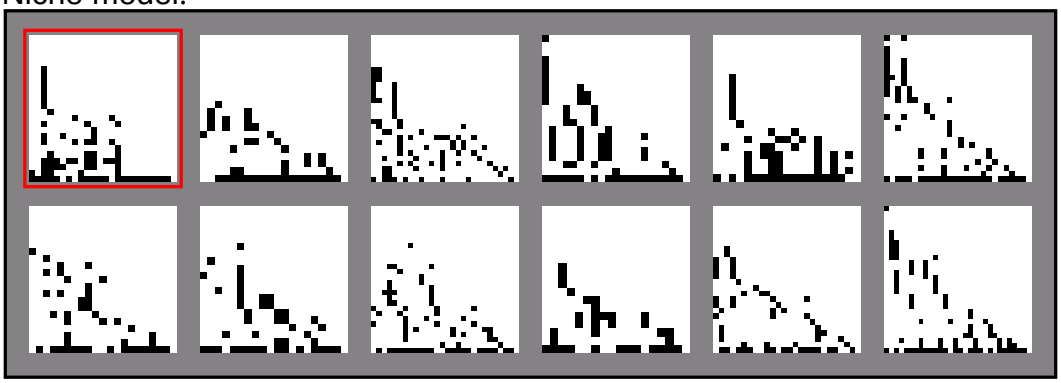


Figure 15: Stony Stream

Matching model:



Niche model:



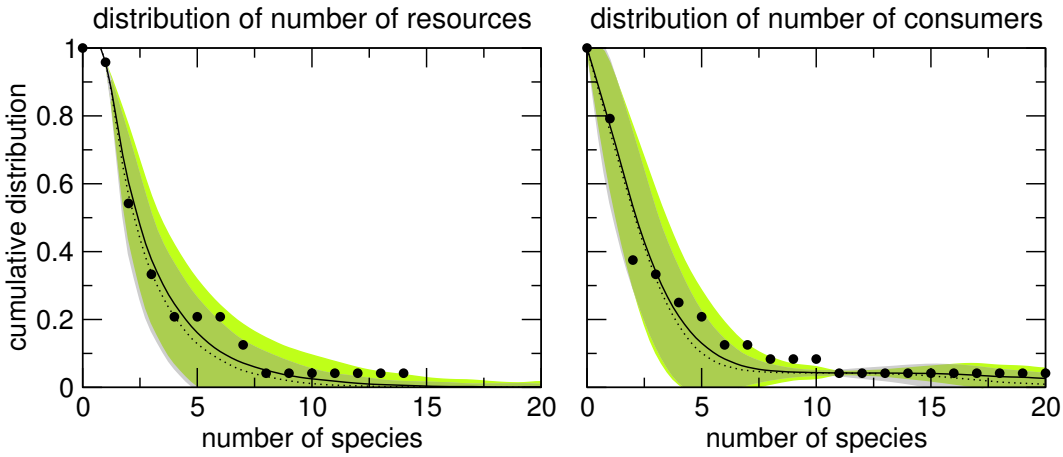
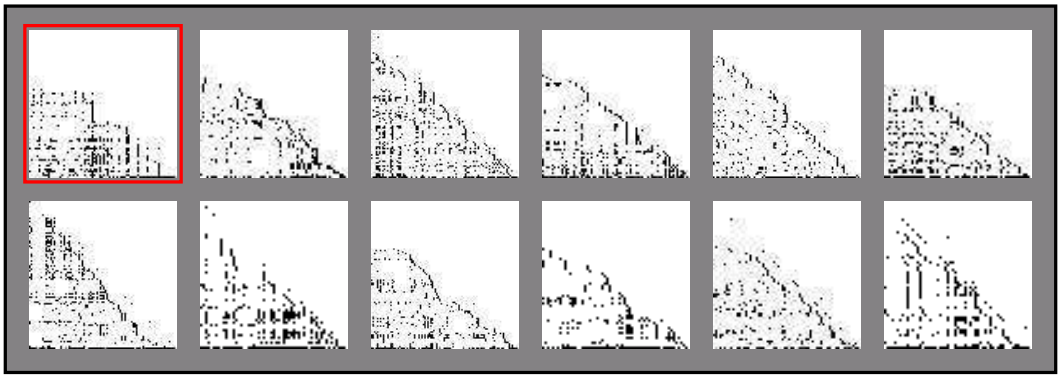
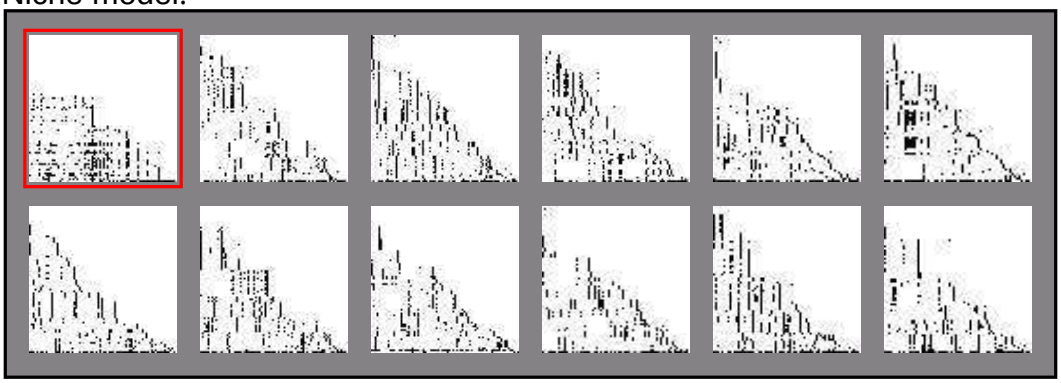


Figure 16: Ythan Estuary 1

Matching model:



Niche model:



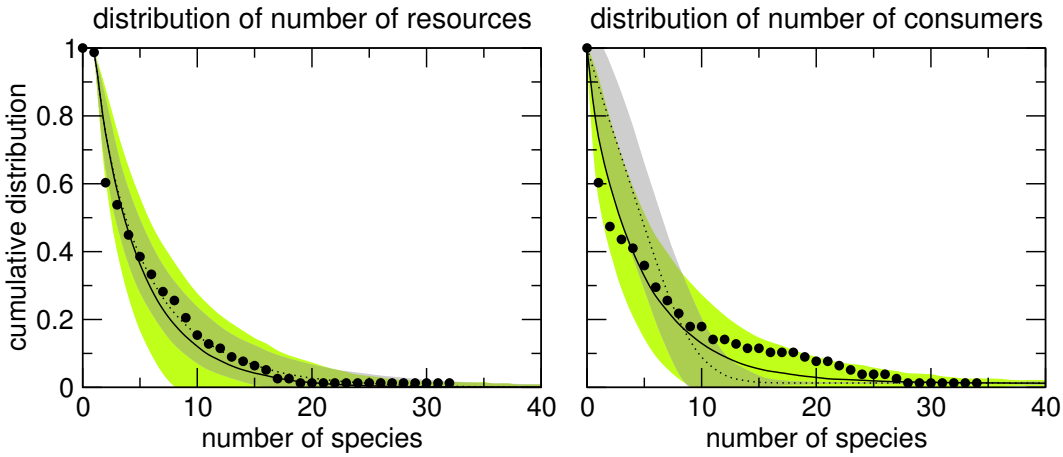
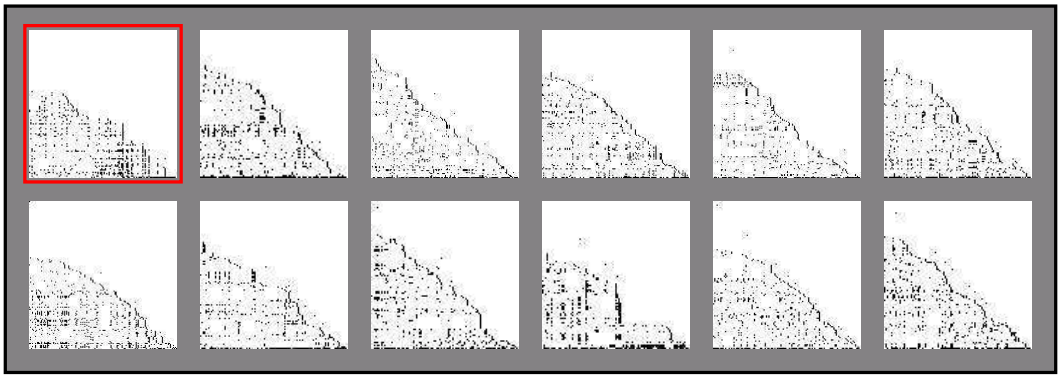


Figure 17: Ythan Estuary 2

Matching model:



Niche model: