Simultaneously estimating food web complexity and structure with uncertainty

1. Food web models are sets of rules which describe and predict the trophic interactions in a food web and hence can help us infer missing interactions among the organisms. One food web model is the allometric diet breadth model (ADBM), which is based on the foraging theory, where the foraging parameters are allometrically scaled to the body sizes of predators and preys. In the study by Petchey et al. (2008), the parameterisation of the ADBM had two limitations: (a) it resulted in only point estimates of the model parameters, and (b) the food web connectance was not estimated, rather a parameter value was fixed to ensure that the predicted connectance was equal to the observed connectance.
2. The novelty of our approach is (a) we consider multiple predictions from the ADBM by parameterising it with approximate Bayesian computation, resulting in estimates of parameter distributions, rather than point estimates. (b) We do not fix connectance to the observed value, but rather let it emerge from the parameterisation. This is acheived by measuring model fit using the true skill statistic, which takes into account prediction of both the presences and absences of links.
3. Employing the new parameterisation method for 16 observed food webs from a wide variety of ecosystems resulted in connectance being consistently overestimated, and structural properties (such as nestedness) being more accurately predicted than in Petchey et al. (2008), hence improving the model predictions. For some food webs, considerable variation in estimated parameter distributions resulting in considerable variation in predicted food web structure.
4. We conclude that the observed food web data likely misses some of the trophic links that can occur, and that the ADBM also predicts some links that cannot occur. The latter could be addressed by accounting in the ADBM for traits other than body size. Further work could also address the significance of the uncertainty in parameter estimates for predicted food web responses to environmental change.

# 1 Introduction

Knowledge about the trophic interactions among the organisms in a community is crucial for understanding the structure and dynamics of ecological communities and for predicting their response to environmental change (Dunne, Williams, and Martinez 2002; Tylianakis and Binzer 2014; O’Connor et al. 2009; Bergamino, Lercari, and Defeo 2011; Krause et al. 2003; Lurgi, L’opez, and Montoya 2012; Morris, Sinclair, and Burwell 2015). This network of trophic interactions is often referred to as a food web. It is important to know the food web structure to answer key ecological questions: which species are more vulnerable to environmental change such as temperature (Petchey et al. 1999); how robust a food web is to extinctions (Dunne, Williams, and Martinez 2002); and how a food web reacts if the top predators are removed (Knight et al. 2005)?

The information about trophic interactions that is used to infer a food web can come from multiple sources, e.g. gut contents (Peralta-Maraver, L’opez-Rodríguez, and de Figueroa 2017), stable isotope composition of tissues (Layman et al. 2007), and experimentation (Warren 1989). Sometimes, the methods used to infer the interactions lead to considerable uncertainty in the constructed food web. E.g. in gut content analysis of some fish predators, there can be tissues which are not identifiable and cannot be assigned to a specific prey item with certainty (Baker, Buckland, and Sheaves 2014). With stable isotope ratios of tissues, uncertainty can be due to factors such as variability in the isotopic fractionation values across multiple combinations of diets and tissues/species, unquantified temporal or spatial variation in prey isotopic values, and variation caused by routing of particular dietary nutrients into particular tissues (Crawford, Mcdonald, and Bearhop 2008). Furthermore, complete recording of all interactions usually requires a large sampling effort even at small spatial and temporal scales (Hobson, Piatt, and Pitocchelli 1994). Food web structure is very difficult to record at larger spatial and temporal scales without losing considerable resolution (spatial, temporal, and taxonomic) (Gravel et al. 2013; Martinez 1991; Jord’an and Osv’ath 2009). Less than complete sampling of interactions can result in no interaction being observed between a pair of individuals that in fact do interact, which results in missing links in a food web. Due to under-sampling, food webs are often poorly understood, which may hinder further advances in the field (Martinez et al. 1999).

When interactions cannot be well observed, and hence well-documented food webs are difficult to obtain, one solution is to use models to predict species interactions (Tamaddoni-Nezhad et al. 2013; Gravel et al. 2013; Petchey et al. 2008; Allesina, Alonso, and Pascual 2008; Cohen, Newman, and Steele 1985). In a sense, the food web model is used to predict missing information about species interactions. For example, Petchey et al. (2008) showed how a model of species interactions (and therefore food web structure) could be parameterised from data on the known presence and absence of trophic interactions. The model and its parameter values encode the rules for occurrence or absence of species interactions and thereby, food web structure. Observed data can be used to select and parameterise the model. Other studies, such as Tamaddoni-Nezhad et al. (2013) used large agricultural datasets and logic-based machine learning and text mining to assign interactions between nodes to construct food webs automatically. Gravel et al. (2013) have developed a method inspired by the niche model of food web structure that used the statistical relationship between predator and prey body size to infer the food web matrix. Food web models have also been used to do ecological forecasting. E.g. Lindegren et al. (2010) used a stochastic food web model driven by regional climate scenarios to produce quantitative forecasts of cod dynamics in the twenty-first century; Hattab et al. (2016) forecasted the potential impacts of climate change on the local food web structure of the highly threatened Gulf of Gabes ecosystem located in the south of the Mediterranean Sea. Hence, food web models have an important role in filling gaps in knowledge about species interactions, including predicting future changes in food web structure.

The allometric diet breadth model (ADBM) was the first model able to predict food web complexity and structure (Beckerman, Petchey, and Warren 2006; Petchey et al. 2008). It used optimal foraging theory, specifically the contingency model (MacArthur and Pianka 1966) to predict which set of the available prey species would be consumed by a predator. This set is the prey species that maximises the energy intake rate of the predator. To do this, the model requires the foraging related traits of species, such as energy content of a potential prey item, the rate of space clearance (also known as attack rate), the density of prey items, and handling time (the amount of time required to handle food items). The model is termed “allometric” because each of these quantities is derived from the body size of the prey and predator using several allometric relationships. The ADBM has also been used to investigate the effect of temperature on an observed food web structure (O’Gorman et al. 2019). A comparison of the ADBM with other food web models has been discussed in Petchey et al. (2011).

The ADBM had variable success in explaining the structure of 15 different food webs, with the proportion of links correctly predicted ranging from 5 % to 65 % (Table ). Coachella valley food web was the best predicted in terms of the proportion of links correct with 65 % of the links correctly predicted by the ADBM. The prediction of the proportion of links correct was poor for Grasslands with 7 % of the links correctly predicted. It was clear that when trophic interactions were more strongly dependent on size, the model correctly predicted a greater proportion of links. Indeed, constructing a food web based only on body size (i.e. ignoring taxonomy) resulted in almost twice the number of links, i.e. 83 % correctly predicted, as when taxonomy used (Woodward et al. 2010).

Although Petchey et al. (2008) demonstrated that foraging theory could predict food web structure, their implementation of the ADBM included at least two limitations. In Petchey et al. (2008), the parameterisation method resulted in estimates of the parameters with no uncertainty. Rather, a single set of parameter values that maximised the explanatory power was selected. In other words, the parameterisation method led to point estimates of the parameters that predicted a single food web structure (because the ADBM is purely deterministic). Moreover, the best prediction is not perfect: the predicted food web was not exactly the same as the observed one. In a sense then, the parameterisation method used in Petchey et al. (2008) was akin to estimating the intercept and slope of a regression line, but not any uncertainty in those parameters. Given that uncertainty is an essential dimension in ecological models, and in predictions about the future state of ecological communities (Petchey et al. 2015; Carpenter 2016), this limitation is important.

The second limitation concerned the estimation of the connectance of the food web, that is the number of realised trophic links divided by the total number of possible trophic links. Although the ADBM can in principle predict connectance, Petchey et al. (2008) did not allow it to do so. Instead, they set the value of relevant parameters in the model to ensure the predicted connectance was equal to the observed connectance. Thus, one can say that the ADBM was not previously used to simultaneously predict complexity and structure of food webs. Moreover, fixing predicted connectance to be equal to observed connectance does not allow any accounting for the possibility that the observed connectance was imperfectly measured. Indeed, if relatively little effort was used to observe the trophic links in a community, the observed connectance will likely be lower than if all trophic links had been observed. Since connectance is an important driver for the stability and dynamics of a food web (May 1972), this limitation needs to be overcome.

In this article we report on how we remove these two limitations, and what we learned when we did so. We removed the first limitation by applying a Bayesian parameterisation approach known as approximate Bayesian computation (ABC). The approach originated from population genetics and has been used in many fields such as systems biology (Toni et al. 2009), ecology (Jabot and Chave 2009), epidemiology (Shriner et al. 2006) and ecological networks (Ibanez 2012; Poisot and Stouffer 2016). An advantage of ABC is that it does not require a likelihood function. Considering that ADBM is a complex deterministic model where the likelihood can not be explicitly evaluated, ABC is a good choice of parameterisation.

We removed the second limitation by allowing estimation of number of links as well as arrangement of links. To accomplish this, we measured model fit by using the true skill statistic, which takes into account both the number of presences and absences of links correctly predicted. High values of the true skill statistic occur when both the predicted arrangement of links and the predicted number of links are close to the observed arrangement and number of links, respectively.

# 2 Materials and Methods

In the upcoming sections, we present a detailed account of the application of the ABC to parameterise the ADBM, along with the description of the ADBM and the food web data used. We explain the rejection Monte Carlo ABC method in the main text, and Markov chain Monte Carlo ABC and sequential Monte Carlo ABC methods in the Supplementary information (hereafter SI) Section S1 (hereafter SI-S1). We computed an accuracy measure known as true skill statistic to assess the ADBM’s predictions and also calculated different food web properties to compare these predictions across food webs.

## 2.1 Allometric Diet Breadth Model (ADBM)

The allometric diet breadth model (ADBM) is based on optimal foraging theory (specifically the contingency foraging model) (MacArthur and Pianka 1966). It predicts the set of prey species a consumer should feed upon in order to maximises its rate of energy intake (Petchey et al. 2008). The species in this set are assumed to be the ones the predator has a trophic link with. In order to make these predictions, the model assumes that a foraging predator is in one of two exclusive states: searching for prey or handling a prey item. The model requires four variables for each potential predator-prey interaction:

* The energy content of the resources (only prey specific) (energy).
* The handling times , which is the time not spent searching caused by consuming a prey item (prey and predator specific) (time).
* The space clearance rates (also known as the attack rate; prey and predator specific) (area or volume per time).
* The prey densities (only prey specific) (individuals per area or volume).

The term “Allometric” in the ADBM refers to the use of four allometric relationships, one for each of these four variables, including the body size of the predator and prey (Table ). With these four allometric relationships, and the body size of each of the species in a community, we can predict the four variables (energy, handling time, space clearance rate, and prey density), and then use the contingency foraging model to predict diets.

Each of the four allometric equations has parameters: a constant and/or at least one exponent (Table ). It is the value of some of these parameters that can be estimated in order to have the predicted food web structure match (as closely as possible) the structure of an observed food web. This is akin to choosing values of slope and intercept of a linear regression that maximises the fit of the regression line to the observed data.

Because some of the allometric constants and exponents are known, and because others are redundant with respect to each other (see Table for details), we are interested in estimating only the following parameters: , , and in the model (Table ).

## 2.2 Observed food web data

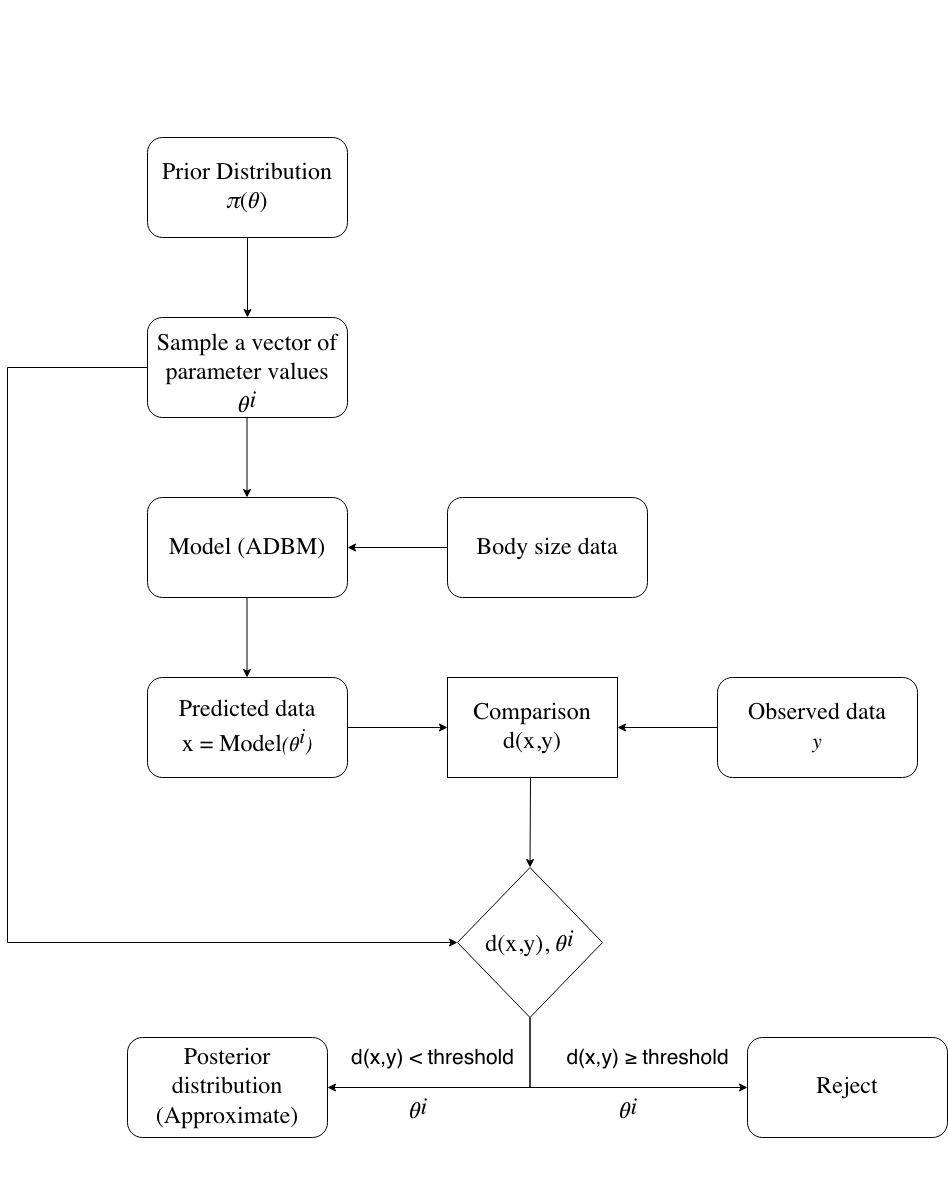
The observed food webs that we fit the ADBM to belong to different types of ecosystems such as marine, freshwater and terrestrial (Table ). The observed connectance of these food webs varies from 0.02 to 0.34 and the number of species varies from 19 to 158. These food webs contain primary producers, herbivores, carnivores, parasites, and parasitoids. They also contain various types of feeding interactions, including predation, herbivory, bacterivory, parasitism, pathogenic, and parasitoid.

The goodness of ADBM’s predictions depends on the types of interactions in these food webs (Petchey et al. 2008). This is because some of the interactions are more size structured than others. Predacious and aquatic herbivore interactions were predicted better than parasitoid and herbivory ones (Petchey et al. 2008). Moreover, the ADBM’s predictions can be better when the species (taxonomic) are replaced by size classes (ignoring taxonomy) (Woodward et al. 2010).

All of the food webs except one (Broadstone Stream) was available only at the species level, i.e. with information about interactions between species and the body size of species. In fact, we use the term “species” here and through this article to indicate a “node” in a food web in which nodes are connected by trophic interactions, and nodes are a collection of individuals that share links. These species/nodes are not always taxonomic, but rather can be broader taxonomic ranks. In contrast, the data for Broadstone Stream was of interactions between individuals, and also contained the body size of the individuals. Thus, the Broadstone Stream food web can be created by aggregating in either of two ways: by taxonomy or by size (Woodward et al. 2010). The ADBM can predict the food web irrespective of the aggregation method, and aggregation by size (while ignoring taxonomy) can lead to much higher explanatory power (i.e. match between observed and predicted food web structure) (83% of links correctly predicted) than with taxonomic aggregation (40% of links correctly predicted) (Woodward et al. 2010).

## 2.3 Parameter estimation: Approximate Bayesian Computation

We used approximate Bayesian computation (ABC) to find sets of parameter values that resulted in predicted food webs that were close in structure to the observation. ABC is an approach that does not require a likelihood function. Instead, there is a distance function that measures the distance between a model’s prediction and the observed data. The approximation of the likelihood depends on the ABC method used, which is further discussed in the upcoming sections. The model parameter values are sampled from a prior distribution. The accepted parameter values form an approximate posterior distribution for the model parameter. We implemented three ABC methods, namely rejection Monte Carlo (Fig. ), Markov chain Monte Carlo, and sequential Monte Carlo to parameterise the ADBM. Since all of these three methods produced very similar results we only include the simplest (rejection) in the main text.



Flowchart of approximate Bayesian computation (rejection) method implemented to parameterise the ADBM.

### 2.3.1 Prior distribution

The prior distribution for and were chosen to be uniform distributions. The range of distribution was from -1.5 to 1.5 and 0 to 3 for and respectively, informed by the estimated in Rall et al. (2012). However, we chose a prior range specific to food webs for the parameter because body size vary greatly among the species in the observed food webs. For example: in the Benguela Pelagic food web, the body sizes of species range from the order of gm to gm. Hence, the range of prey-predator ratio was from the order of to . To take this into account, we took the prior of from a uniform distribution ranging from to . In the case of parameter , we chose the prior of to be a uniform distribution. However, the prior range varied between food webs. For example the prior range for Benguela Pelagic was chosen to be -12 to 10. The method for choosing the specific range of the parameter is detailed in the SI-S5.

### 2.3.2 Comparison of observed and predicted

The difference between the model’s prediction and the observed data (e.g. the sum of squared residuals is such a distance in linear regression) is quantified by a distance measure. The distance is lower when there is a closer match between the model’s prediction and the observation. A perfect match would result in zero distance.

The magnitude of the distance is used to decide on the acceptance or rejection of a set of parameter values. An accepted set of parameter values contributes to the posterior distribution, rejected ones do not. This makes the distance measure one of the important features of ABC. A threshold distance is chosen, and if the distance for a particular set of parameter values is less than the threshold, then that set of parameter values contributes to the posterior distribution. Otherwise, when the distance is greater than the threshold, the parameter values do not contribute to the posterior. Hence, the magnitude of the distance threshold determines the proportion of a model’s parameters that are accepted. A higher threshold causes a high proportion of acceptances but less accuracy (some parameter sets that result in predictions quite unlike the observed data are accepted). Below, we first describe and justify our choice of distance measure, and then our choice of threshold.

In the study of Petchey et al. (2008) the measure of distance was equivalent to , where is the number of observed links that were predicted (the number of true positives) and is the number of observed links that were not predicted (the number of false negatives). A distance of 0 indicates that all observed links were correctly predicted. One way for the ADBM to achieve this is to predict that every species has a trophic link with every other species (including itself) – a fully connected food web with connectance of 1. Petchey et al. (2008) prevented this by constraining the number of predicted links to be equal to the number of observed links (i.e. they fixed the model connectance to be the same as the observed connectance, as previously mentioned). In this study, we wish to relax this constraint, and instead have the number of links be estimated, as well as the arrangement of links. The first step of doing so was choosing an appropriate distance measure.

The distance measure used in this study was 1 minus the true skill statistic: . This distance ranges from 0 to 2.

TSS is defined as:

where is the number of observed links that are predicted by the model (true positives), is the number of observed absences of links that are correctly predicted (true negatives), is the number of false positives, and is the number of false negatives.

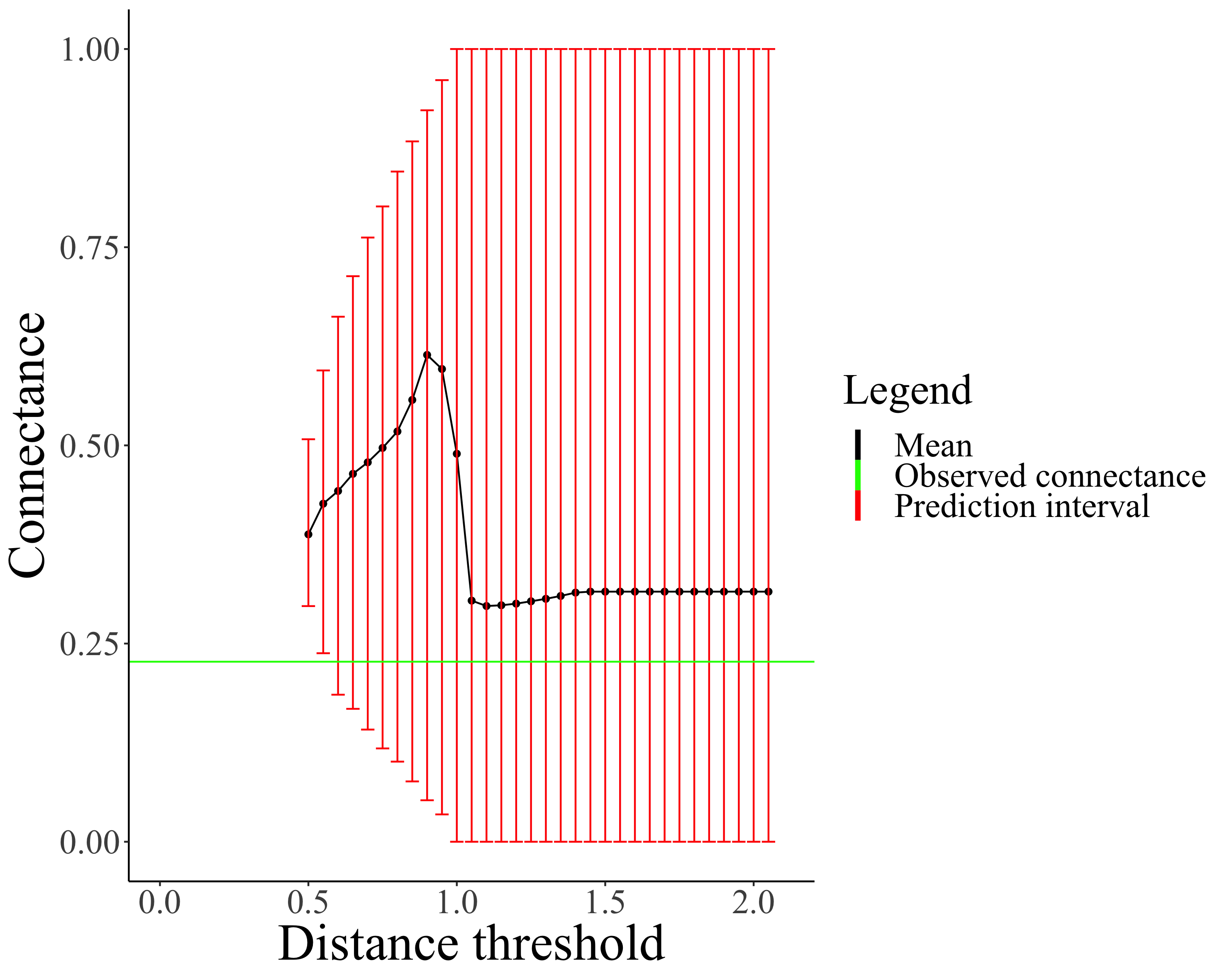
The ranges from to , where +1 indicates a perfect prediction. A value of zero or less indicates a performance no better than random.

The inclusion of true and false negatives in the distance measure means that the best theoretically possible prediction (smallest distance) is a unique prediction, and specifically the one in which the predicted presence and absence of links matches exactly with the observed presence and absence of links.

We now describe how we chose the distance threshold. Because food web dynamics and stability are strongly dependent on connectance (May 1972), we set the distance threshold such that the model had a reasonable chance of predicting the observed value of connectance.

To do this, we examined how the predicted connectance varied with the tolerance threshold. An example of this relationship is given in Fig. for the Benguela Pelagic food web. We then chose the minimum threshold value that gave a range of predicted connectance that contained the observed connectance.

Furthermore, it is useful to note that in Fig. there are no connectance values below a distance threshold value of less than 0.5. This is because, for this particular food web, there were no sets of parameter values that could achieve a better model fit than is indicated by . I.e. it is simply impossible for the ADBM to make better predictions than this. One reason for this is that the ADBM (when body size is the only trait) can only predict contiguous diets in trait space, whereas the observed data contains gaps in the diet.



The prediction interval of the predicted connectance increases with increasing distance threshold for Benguela Pelagic food web. The green line and black line represent the observed connectance and mean of predicted connectance respectively.

### 2.3.3 The Rejection ABC method

In the rejection ABC method, a set of parameter values are sampled from the prior distributions. This set of parameter values is either accepted, and thereby added to the posterior distribution of the parameter values, or it is rejected (based on if the distance 1 - is less than or greater than the threshold distance, as mentioned above). This process is repeated until there are enough acceptances to give stable (approximate) posterior distributions. In addition, we used a kernel function that assigns weight to each set of parameter values, where the weight is inversely proportional to the distance (1 - ).

In the upcoming section, we further detail the rejection ABC method.

*Properties:*

* A prior distribution : is the uniform distribution for parameters
* A model prediction : ADBM. This is a predicted food web, , given by a particular set of parameter values . Hence,
* A summary statistic : is the predation matrix predicted by the ADBM.

where is the distance threshold

* A distance function :
* An observed food web , in the form of a predation matrix containing zeros and ones.

*Sampling:*

for

* Draw a set of parameter values from the prior distribution .
* Compute the model result
* Compute and
* Accept or reject the parameter set probabilistically:
  + Assign a probability to as per the kernel ; , where is the distance evaluated in the previous step.
  + Compute
  + If , then accept and

*Output:*

An approximate joint posterior distribution using the accepted .

## 2.4 Assessment of model fit

Accuracy refers to how close is the model prediction to the observation. As the ADBM’s prediction is a predation matrix that consists of the presence and absence of links, comparing how close the prediction is to the observation is not straightforward as comparing two numerical values. Hence, to take into account the true and false predictions of both the presence and absence of links, we defined the accuracy of the ADBM using true skill statistics which is already defined above.

We also examined how closely structural properties of the predicted food web matched those of the observed food webs. We evaluated properties such as proportion of basal species, proportion of intermediate species, proportion of top species, proportion of herbivores, mean omnivory, clustering coefficient, standard deviation of generalism, standard deviation of vulnerability, diet similarity and mean path length. We did not compute mean trophic level and maximum trophic level because their computation did not converge in the R *cheddar* package (Hudson et al. 2013) for all the food webs.

We investigated the performance of the ADBM parameterised with the ABC by computing standardised error of the food web properties, where the standardised error is the absolute raw error (the difference between observed and predicted value) divided by the maximum absolute raw error for that property. We did not calculate the standardised error for mean omnivory because it had some NA values due to non-convergence of computation in the R *cheddar* package (Hudson et al. 2013) for all the food webs.

# 3 Results

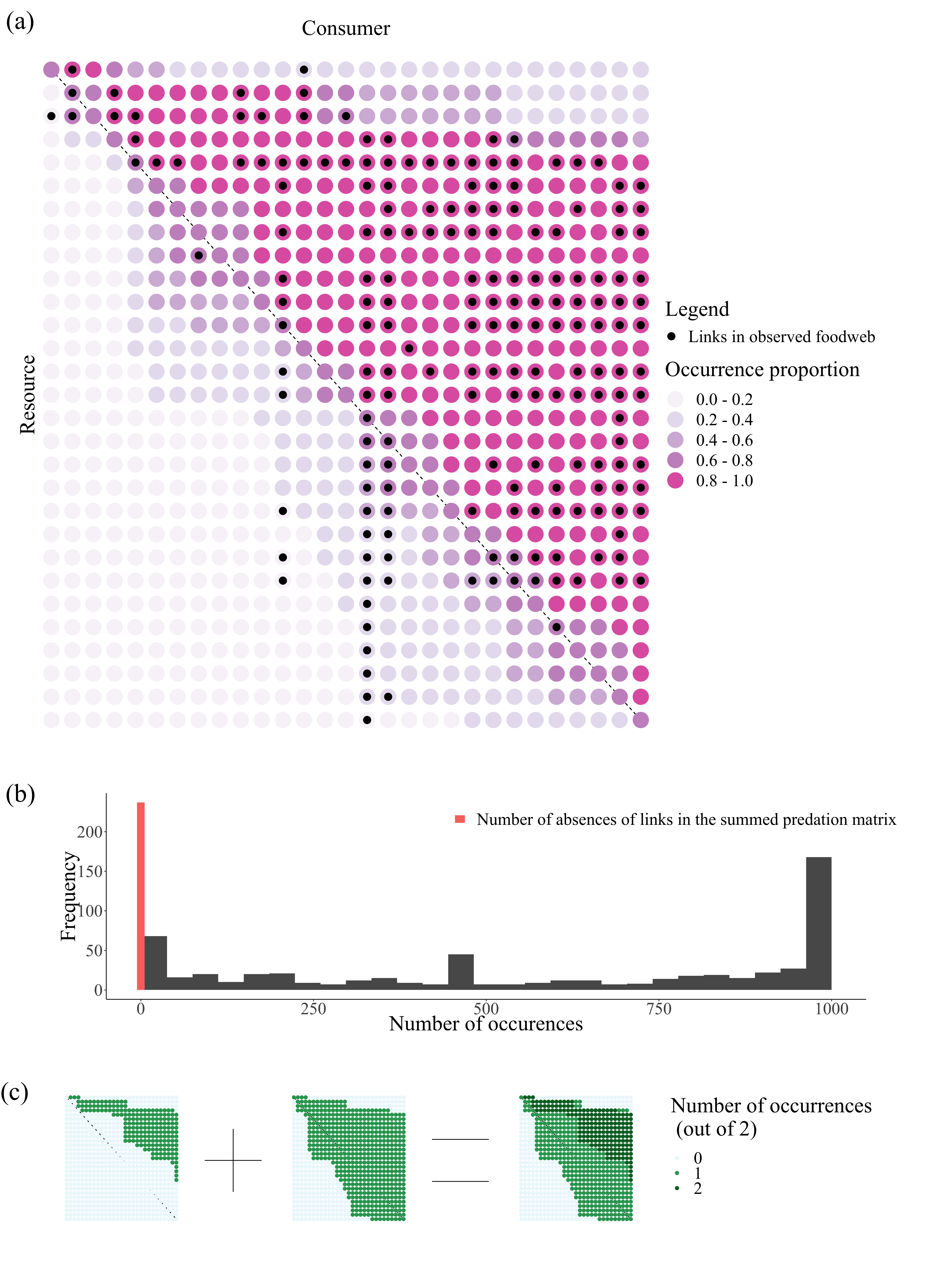
We first present the results for the Benguela food web, as an example of the model outcomes (e.g. predicted food web structure, variation in predicted food web structure, and posterior parameter distributions). We chose this food web as it was relatively well explained using the method of Petchey et al (2008). Results for all other food webs are presented in the SI. We then present a comparison across all the empirical food webs of model outcomes between those of Petchey et al (2008) and our current work, e.g, we compare the true skill statistic of the two approaches and compare some food web properties, such as proportions of basal, intermediate, and top species.

## 3.1 Exemplary results (Benguela food web)

The true skill statistic (TSS) of the predicted Benguela Pelagic food web varied between 0.4 and 0.52. This variation in the TSS is represented in terms of predation matrices displayed in Fig. (a), which overlays 1000 independent predation matrices created from the posterior parameter distributions. In all the 1000 independent predation matrices the predicted links are mostly present in the upper triangular portion of the matrix where most of the observed links are also present. (Links in the upper right triangle of the predation matrix are for predators feeding on prey smaller than themselves.)

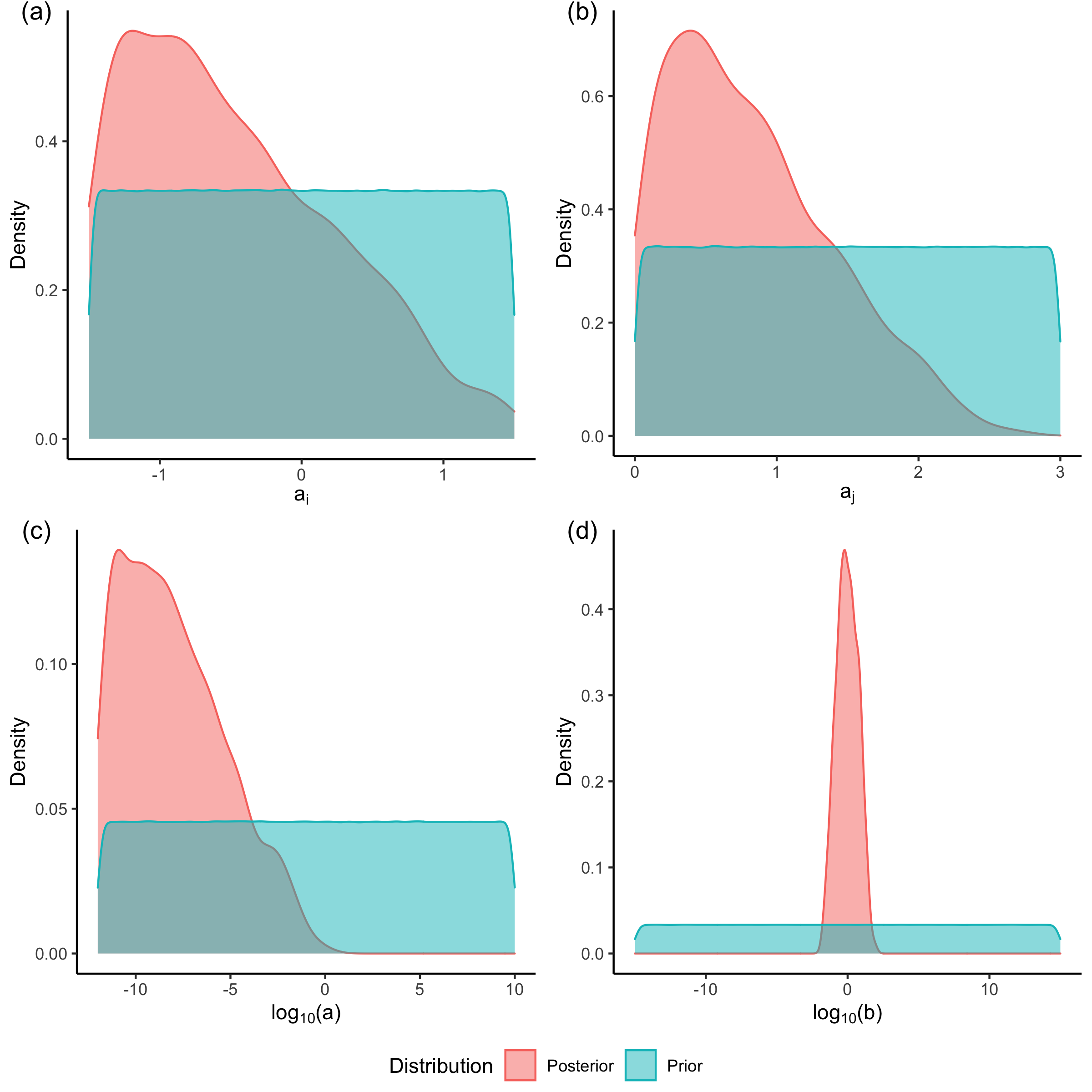
In the 1000 predicted predation matrices, there are sometimes predators that are smaller than their predicted prey (links in the lower left triangle of the predation matrix). This is also portrayed in the marginal distribution of in Fig. (d), as it includes values greater than (). This is relevant as values of make the most profitable prey item equal in size to the predator size. Lower values of make the most profitable prey item smaller than the size of the predator.

There were around 250 potential links in the lower left triangle of the predation matrix that were never predicted in any of the 1000 predicted predation matrix (Fig. (b)). This strongly suggests that the predator-prey size ratio of these links is so small (i.e. very large prey, very small predator) that the links cannot occur, given that the preponderance of observed links is predators consuming prey small than themselves.



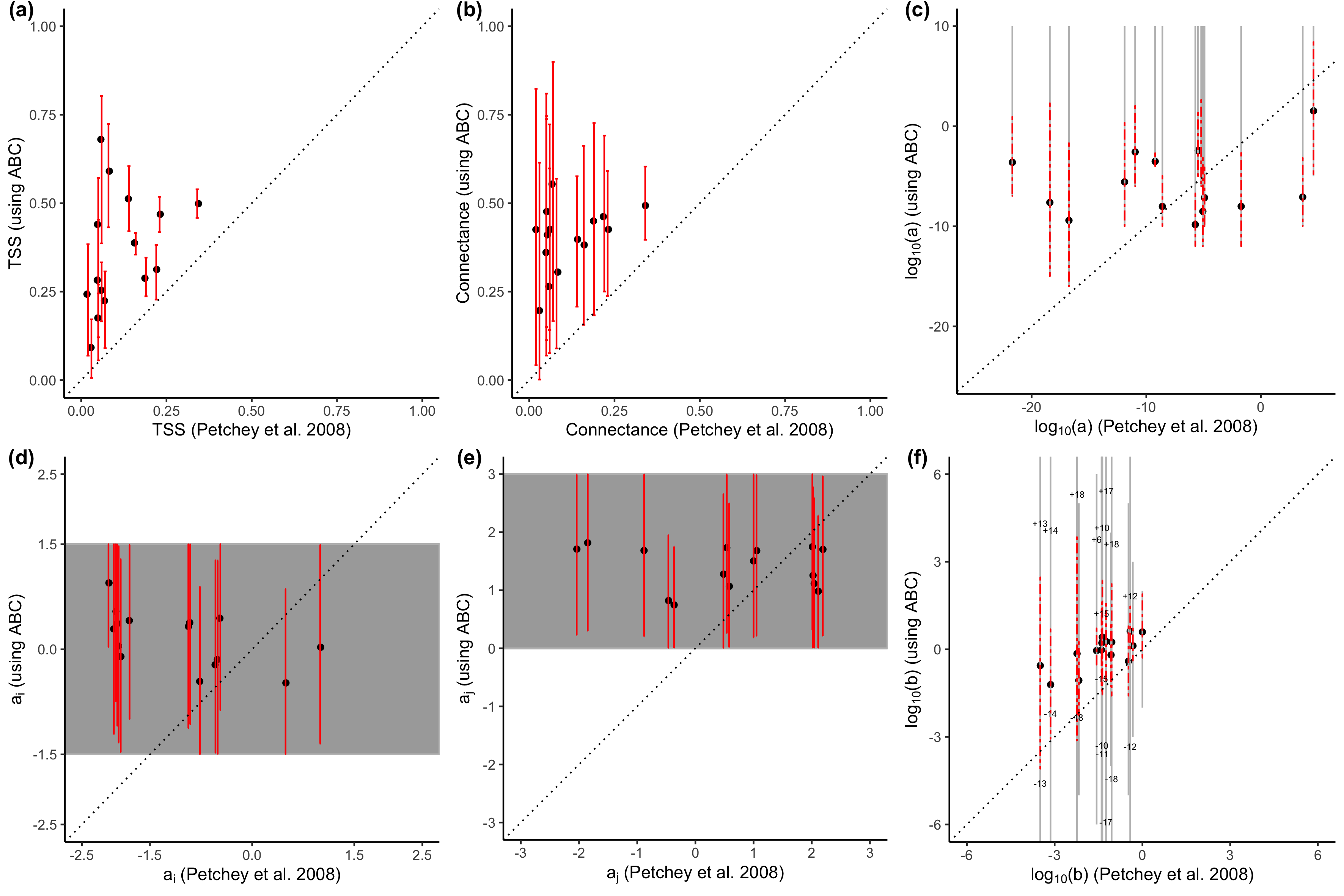
(a) Observed and predicted predation matrices for Benguela Pelagic food web. Body size increases from left to right and top to bottom along the predation matrix. Black circles show where there is an observed trophic link. The intensity of the pink circles shows the proportion of 1000 predicted food webs that had a trophic link between the corresponding species. This type of overlay is shown for two examples predicted in panel (c). (b) The histogram of the number of times a link was predicted across the 1000 independently predicted food webs. There were 841 species pairs in this food web. About 150 of these were predicted to have a trophic link in all 1000 predicted predation matrices. The red bar shows the number of pairs of species for which a trophic link was never predicted. (c) Two predicted predation matrices for Benguela Pelagic food web corresponding to the minimum and the maximum value of estimated , and their sum.

The marginal posterior of parameter in the Benguela Pelagic food web was more constrained than the other three allometric parameters (Fig. ) i.e. the posterior range was the narrowest.



Marginal prior and marginal posterior distribution of the ADBM parameters for the Benguela Pelagic food web estimated using rejection ABC.

## 3.2 Comparison with Petchey et al. (2008)



TSS (a), connectance (b) and ADBM parameters (c, d, e, f) computed using the ABC method compared with the corresponding point estimates from Petchey et al (2008). The red lines are the 95% credible/prediction intervals and the black filled circles represent the corresponding means. The grey region represents the intervals of the prior distributions for and . The grey lines represent the prior range of the parameters and in the scale. The prior range for the parameter extends above and below the y-axis limits for some food webs and so the values of the limits are shown on the plot. The dashed black lines are the 1:1 relationships for reference.

Across all of the food webs, the mean true skill statistic using the ABC approach was higher than that of the point estimates from Petchey et al. (2008) (hereafter referred as previous study) (Fig. (a)). Furthermore, the 95% predicted interval of the TSS for the ABC approach was above and did not overlap the TSS from Petchey et al. (2008). Across all of the food webs, our present approach led to estimates of connectance greater than the values of connectance from Petchey et al. (2008) (which were then fixed to equal the observed values of connectance).

We did not find a consistent relationship between the parameters estimated using the current approach and those estimated in Petchey et al. (2008) (Fig. (c-f)), except for in the case of parameter . In this case, the mean from the current approach was always higher than the estimates from Petchey et al. (2008) (Fig. (f)) and the 95% credible of the posterior of always includes the estimate from Petchey et al. (2008).

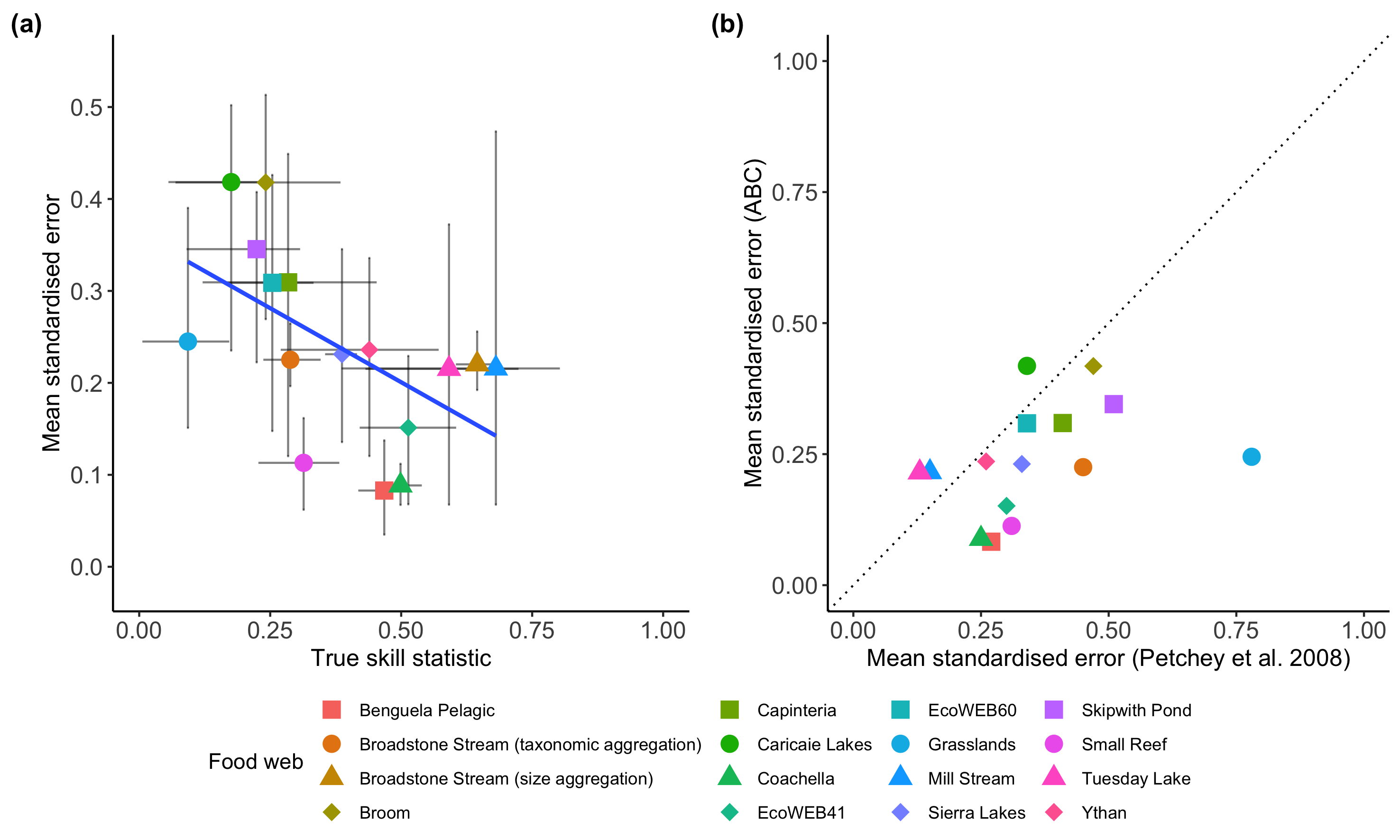
The marginal posterior of parameter was more constrained than the other three allometric parameters, i.e. the posterior range was the narrowest (SI Fig. S21-S36). In most of the food webs, the parameter had a unimodal distribution (SI Fig. S21-S36). EcoWEB60 and Grasslands had a bimodal distribution while Sierra Lakes had three modes.

The structural food web properties proportion of intermediate species, mean omnivory, clustering coefficient, sd of generality, sd of vulnerability, diet similarity and nestedness estimated from the current ABC approach were mostly higher than that from the previous study (SI Fig. S40(b, e-j)). The properties proportion of basal species, proportion of top species, and proportion of herbivores were mostly lower (SI Fig. S40(a, c, d)).

For proportion of intermediate species, mean omnivory, clustering coefficient, sd of generality, sd of vulnerability and nestedness, the real values were mostly within the lower range of the predicted 95% interval. In the case of proportion of basal species, proportion of top species, proportion of herbivores these properties were underestimated when compared to the real values for most of the food webs.

A principal component analysis of variation in the food web structural properties revealed a first PC axis representing on average 62% of the overall variance. This first axis was highly correlated with connectance, with an average Spearman correlation of 0.87.

The ADBM when parameterised with the ABC predicted the structural food web properties (such as proportion of basal species) generally better when the true skill statistics was higher (Fig. (a)) when compared across the 16 food webs. Furthermore, the ABC parameterised ADBM made more accurate predictions of food web properties than the predictions in Petchey et al (2008) (Fig. (b)).



(a) The mean standardised error of the food web properties predicted from the ADBM parameterised using rejection ABC plotted against the mean true skill statistic for each food webs. The vertical and horizontal bars correspond to 95% prediction intervals of the standardised error and true skill statistic respectively. Solid blue line is linear regression through the means (t = -2.44, df = 14, P = 0.028). (b) The mean standardised error computed from the ABC method plotted against the mean standardised error from Petchey et al. (2008). The dashed line is the 1:1 relationship for reference.

Within each food web, we found various relationships between the standardised error and true skill statistic (SI Fig. S41). E.g. For Skipwith Pond food web (SI Fig. S41(l)), high values of TSS were associated with high error, whereas the opposite was true for other food webs, such as Broadstone Stream (SI Fig. S41(b, p)). Other food webs showed more complex relationships.

# 4 Discussion

The ABC parameterisation method employed here greatly improves on the basic parameterisation methods applied in Petchey et al. (2008). The ABC method provides uncertainty in parameter estimates, and thereby a range of predicted food webs (Fig. (c-f)). It also allowed us to estimate parameters that were fixed by Petchey et al. (2008), and thereby also predict connectance (rather than fix it to the observed value) (Fig. (b)). Including uncertainty and predicting connectance are, we believe, significant advances. They allow, for example, predictions of changes of food web structure caused by environmental change that include uncertainty in the predicted food web structure; including uncertainty in such predictions is critical (Petchey et al. 2015; Cressie et al. 2009; Lindegren et al. 2010). A future endeavor will be to partition the contribution of different sources of uncertainty (e.g. incomplete sampling, model deficiencies), in order to know where to attempt to make improvements aimed at reducing uncertainty. Another avenue for future research is to investigate the functional and dynamical significance of the uncertainty in the predicted food web structure. Below we discuss some other lessons from the results of our study, and also expand on these opportunities and priorities for future research.

## 4.1 Connectance

In all cases, the predicted connectance was greater than the observed connectance (Fig. (b)). Why did this occur? First, it is important to recognise that the ADBM can only predict diets that are contiguous with respect to the size of prey. I.e. it cannot predict that a predator will consume prey of size 1 and 3, and not prey of size 2. (It can, however, predict such patterns if a trait other than size, and not perfectly correlated with size, influences foraging parameters (Petchey et al. 2008). Second, it is important to note that observed diets were not contiguous with respect to prey size, i.e. observed diets contain holes when prey are ordered by their size. Given these features, and that the model is attempting to maximise the coincidence of predicted and observed link presence and absence (i.e. the true skill statistic), the estimation process will result in a greater number of predicted links than observed.

These findings raise the question of whether the model is wrong or whether the observed data is wrong. Likely both are wrong to some extent. We expect that the observed data contains some absences of links that would actually occur in reality. Sampling effort that is too low to stand much chance of observing rare feeding interactions could result in this. In this case, the model could be correctly predicting a link that was not yet observed (model correct, data incorrect). More intensive and more complete sampling of links in food webs has for many years been recognised as important, due to the potential for low sampling effort to influence perceived food web structure (Martinez et al. 1999).

We also expect there are cases where the model is wrongly predicting that a feeding link will occur, when there is no possibility that one would in reality occur. This could happen when a trait other than, or in addition to, prey size is influential. For example, a particular prey species may have a defensive trait that means it takes longer to consume than one of the same size that is undefended. An important development here would be to incorporate traits other than body size in the ADBM to allow discontiguous diets along the size axis. Furthermore, in its current form, the ADBM is a biology-only model; it does not include an observation process, though this could be added, if there were evidence that sampling was incomplete. This could allow the model to predict the absence of a link due to incomplete observation.

It would be interesting to take a very well sampled food web, and test if the ABC parameter estimation applied to a subset of the observed links (i.e. simulating a poorly sampled food web) does predict the connectance of the very well sampled food web. If so, this would indicate the potential to compensate for under-sampling with an appropriate food web model and estimation procedure.

We were somewhat surprised to observe ABC parameterisation have greater prediction accuracy of structural features of the food webs (Fig. ). We expected the overestimation of connectance to cause worse prediction accuracy. Indeed, the greater the error in connectance, the greater the error in predicted structure. However, the ABC parameterisation had higher true skill statistic, suggesting that the effect of this on accuracy outweighed the effect of greater error in connectance. That is, the overestimation of connectance was not so critical for accurate predictions of structure, and a higher TSS was more influential.

## 4.2 Observing and predicting link absences

Our parameterisation approach attempted to maximise the true skill statistic (the coincidence of predicted and observed link presences, and the coincidence of predicted and observed link absences). The TSS assigns equal importance to the collection of presence and absence of observed links. And the amount of weight to a single presence or absence of observed link is dependent on the connectance of the food web. If the connectance is less than 0.5, the TSS assigns more weight to a presence of link as compared to an absence of link and vice versa.

Because the connectance of the observed food webs is less than 0.5 (Table ), the TSS assigned more importance to a single presence of link when compared to a single absence of link. And assigning more weight to a single presence of link when compare to a single absence of link in our study is logical because the chance of a recorded presence is a truly a presence is likely to be greater than the chance that a recorded absence is truly an absence. This is because observation of a single feeding interaction is sufficient to record the presence of a link. However, this is not true for the absence of links: one observation of a predator not consuming a prey does not mean that it, or another predator of the same type, never would. That is, if we observe no interaction between two species during the sampling period, we conclude that there is an absence of link.

A possible improvement of our estimation procedure could quantify the uncertainty in the recorded absence of links and to include this uncertainty in the parameterisation method. This might be accomplished by assigning different weight/importance to true positives, true negatives, false positives and false negatives calculated from empirical studies which might be specific to a food web. A different option would be to add an observation process to the model, such that the biological part of the model can predict that a feeding link is possible, but then the observation process leads to that link not being observed.

## 4.3 Importance of different allometric parameters

In Petchey et al. (2008), the parameter played a major role in maintaining the maximum predictive power of the ADBM. Indeed, they found that estimating only, and not estimating either of or led to only a slight decrease in model performance, and that estimating only and led to no decrease in model performance relative to when all three parameters were estimated.

Likewise, we found that the posterior distribution of the parameter was the most constrained among all the parameters (Fig. ). The parameter defines the range of prey body size which has a finite handling time, and the prey size with the highest energetic profitability. As the parameter relates to the prey-predator body size ratio, the constrained posterior of (Fig. (d)) tells us the importance of the ratio of body size of prey and predator in determining the food web structure in the ADBM.

The marginal posterior of parameter was right-skewed (Fig. (c)). This may be because the ABC parameterisation overestimates the connectance, which means that lower values of are preferred over higher values of (a lower value of leads to a lower space clearance/attack rate, and a lower space clearance rate results in a higher connectance).

## 4.4 Improving parameterisations by using other observed data

Information about who eats who can come from multiple sources, such as gut contents of organisms, stable isotope composition of tissues, and experimentation (Peralta-Maraver, L’opez-Rodríguez, and de Figueroa 2017; Layman et al. 2007; Warren 1989). Moreover, experimentation can lead to independent estimates of allometric foraging parameters, such as , , and (Rall et al. 2012). A prospect would be to investigate how diverse data can be used to parameterise the ADBM’s predictions and how uncertainty in these different data influence the ADBM’s predictions using ABC. With the help of appropriate summary statistics in the ABC method, one can answer these questions. E.g. We could use the approximate trophic position inferred from stable isotope ratio data from an individual tissue and gut content data of a predator simultaneously to parameterise the ADBM. In this example, the trophic position and the gut content information would be the summary statistics. Another important question is how the quantity of data affect the ADBM’s predictions. This can help food web researchers decide how much data from a specific source is enough to predict the food web structure, and help further optimise the deployment of limited sampling resources.

Sometimes, only partial food web data is available (Patonai and Jord’an 2017). With the help of summary statistics in the ABC, partial data can be used for inferring these food web structures from the ADBM. E.g. We could use gut content data of only some of the species in a food web to parameterise the ADBM, thereby predicting the food web structure. Summary statistics opens up a broad spectrum of possibilities in parameterising the food web models. There are multiple empirical and theoretical studies on a range of different food web properties of food webs across different ecosystems (Williams and Martinez 2000; Goldwasser and Roughgarden 1993; Martinez 1991). These can conceivably be used in parameterising food web models using ABC to constraint the model predictions.

# 5 References

Allesina, Stefano, David Alonso, and Mercedes Pascual. 2008. “A General Model for Food Web Structure.” *Science* 320 (5876). American Association for the Advancement of Science: 658–61. <https://doi.org/10.1126/science.1156269>.

Baker, Ronald, Amanda Buckland, and Marcus Sheaves. 2014. “Fish Gut Content Analysis: Robust Measures of Diet Composition.” *Fish and Fisheries* 15 (1): 170–77. <https://doi.org/10.1111/faf.12026>.

Beckerman, A. P., O. L. Petchey, and P. H. Warren. 2006. “Foraging Biology Predicts Food Web Complexity.” *Proceedings of the National Academy of Sciences of the United States of America* 103: 13745–9.

Bergamino, Leandro, Diego Lercari, and Omar Defeo. 2011. “Food Web Structure of Sandy Beaches: Temporal and Spatial Variation Using Stable Isotope Analysis.” *Estuarine, Coastal and Shelf Science* 91 (4): 536–43. <https://doi.org/10.1016/j.ecss.2010.12.007>.

Carpenter, Stephen R. 2016. “Ecological Futures: Building an Ecology of the Long Now.” *Ecology*, October, 2069–83. <https://doi.org/10.1890/0012-9658(2002)083[2069:EFBAEO]2.0.CO;2@10.1002/(ISSN)1939-9170.MacArthurAward>.

Cohen, Joel E., C. M. Newman, and John Hyslop Steele. 1985. “A Stochastic Theory of Community Food Webs I. Models and Aggregated Data.” *Proceedings of the Royal Society of London. Series B. Biological Sciences* 224 (1237). Royal Society: 421–48. <https://doi.org/10.1098/rspb.1985.0042>.

Crawford, Kerry, Robbie A. Mcdonald, and Stuart Bearhop. 2008. “Applications of Stable Isotope Techniques to the Ecology of Mammals.” *Mammal Review* 38 (1): 87–107. <https://doi.org/10.1111/j.1365-2907.2008.00120.x>.

Cressie, Noel, Catherine A. Calder, James S. Clark, Jay M. Ver Hoef, and Christopher K. Wikle. 2009. “Accounting for Uncertainty in Ecological Analysis: The Strengths and Limitations of Hierarchical Statistical Modeling.” *Ecological Applications* 19 (3): 553–70. <https://doi.org/10.1890/07-0744.1>.

Dunne, Jennifer A., Richard J. Williams, and Neo D. Martinez. 2002. “Network Structure and Biodiversity Loss in Food Webs: Robustness Increases with Connectance.” *Ecology Letters* 5 (4): 558–67.

Goldwasser, Lloyd, and Jonathan Roughgarden. 1993. “Construction and Analysis of a Large Caribbean Food Web: Ecological Archives E074-001.” *Ecology* 74 (4): 1216–33. <https://doi.org/10.2307/1940492>.

Gravel, Dominique, Timoth’ee Poisot, Camille Albouy, Laure Velez, and David Mouillot. 2013. “Inferring Food Web Structure from Predator-Prey Body Size Relationships.” Edited by Robert Freckleton. *Methods in Ecology and Evolution* 4 (11): 1083–90. <https://doi.org/10.1111/2041-210X.12103>.

Hattab, Tarek, Fabien Leprieur, Frida Ben Rais Lasram, Dominique Gravel, François Le Loc’h, and Camille Albouy. 2016. “Forecasting Fine-Scale Changes in the Food-Web Structure of Coastal Marine Communities Under Climate Change.” *Ecography* 39 (12): 1227–37. <https://doi.org/10.1111/ecog.01937>.

Hobson, Keith A., John F. Piatt, and Jay Pitocchelli. 1994. “Using Stable Isotopes to Determine Seabird Trophic Relationships.” *Journal of Animal Ecology* 63 (4). [Wiley, British Ecological Society]: 786–98. <https://doi.org/10.2307/5256>.

Hudson, Lawrence N., Rob Emerson, Gareth B. Jenkins, Katrin Layer, Mark E. Ledger, Doris E. Pichler, Murray S. A. Thompson, Eoin J. O’Gorman, Guy Woodward, and Daniel C. Reuman. 2013. “Cheddar: Analysis and Visualisation of Ecological Communities in R.” *Methods in Ecology and Evolution* 4 (1): 99–104. <https://doi.org/10.1111/2041-210X.12005>.

Ibanez, S’ebastien. 2012. “Optimizing Size Thresholds in a Plant–Pollinator Interaction Web: Towards a Mechanistic Understanding of Ecological Networks.” *Oecologia* 170 (1): 233–42. <https://doi.org/10.1007/s00442-012-2290-3>.

Jabot, Franck, and J’erôme Chave. 2009. “Inferring the Parameters of the Neutral Theory of Biodiversity Using Phylogenetic Information and Implications for Tropical Forests.” *Ecology Letters* 12 (3): 239–48. <https://doi.org/10.1111/j.1461-0248.2008.01280.x>.

Jord’an, Ferenc, and Györgyi Osv’ath. 2009. “The Sensitivity of Food Web Topology to Temporal Data Aggregation.” *Ecological Modelling* 220 (22): 3141–6. <https://doi.org/10.1016/j.ecolmodel.2009.05.002>.

Knight, Tiffany M., Michael W. McCoy, Jonathan M. Chase, Krista A. McCoy, and Robert D. Holt. 2005. “Trophic Cascades Across Ecosystems.” *Nature* 437 (7060): 880–83. <https://doi.org/10.1038/nature03962>.

Krause, Ann E., Kenneth A. Frank, Doran M. Mason, Robert E. Ulanowicz, and William W. Taylor. 2003. “Compartments Revealed in Food-Web Structure.” *Nature* 426 (6964): 282–85. <https://doi.org/10.1038/nature02115>.

Layman, Craig A., D. Albrey Arrington, Carmen G. Montaña, and David M. Post. 2007. “Can Stable Isotope Ratios Provide for Community-Wide Measures of Trophic Structure?” *Ecology* 88 (1): 42–48. <https://doi.org/10.1890/0012-9658(2007)88[42:CSIRPF]2.0.CO;2>.

Lindegren, Martin, Christian Möllmann, Anders Nielsen, Keith Brander, Brian R. MacKenzie, and Nils Chr. Stenseth. 2010. “Ecological Forecasting Under Climate Change: The Case of Baltic Cod.” *Proceedings of the Royal Society B: Biological Sciences* 277 (1691): 2121–30. <https://doi.org/10.1098/rspb.2010.0353>.

Lurgi, Miguel, Bernat C. L’opez, and Jos’e M. Montoya. 2012. “Climate Change Impacts on Body Size and Food Web Structure on Mountain Ecosystems.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1605): 3050–7. <https://doi.org/10.1098/rstb.2012.0239>.

MacArthur, Robert H., and Eric R. Pianka. 1966. “On Optimal Use of a Patchy Environment.” *The American Naturalist* 100 (916): 603–9.

Martinez, Neo D. 1991. “Artifacts or Attributes? Effects of Resolution on the Little Rock Lake Food Web.” *Ecological Monographs* 61 (4): 367–92. <https://doi.org/10.2307/2937047>.

Martinez, Neo D., Bradford A. Hawkins, Hassan Ali Dawah, and Brian P. Feifarek. 1999. “Effects of Sampling Effort on Characterization of Food-Web Structure.” *Ecology* 80 (3): 1044–55. <https://doi.org/10.1890/0012-9658(1999)080[1044:EOSEOC]2.0.CO;2>.

May, Robert M. 1972. “Will a Large Complex System Be Stable?” *Nature* 238 (5364): 413. <https://doi.org/10.1038/238413a0>.

Morris, Rebecca J., Frazer H. Sinclair, and Chris J. Burwell. 2015. “Food Web Structure Changes with Elevation but Not Rainforest Stratum.” *Ecography* 38 (8): 792–802. <https://doi.org/10.1111/ecog.01078>.

O’Connor, Mary I., Michael F. Piehler, Dina M. Leech, Andrea Anton, and John F. Bruno. 2009. “Warming and Resource Availability Shift Food Web Structure and Metabolism.” Edited by Michel Loreau. *PLoS Biology* 7 (8): e1000178. <https://doi.org/10.1371/journal.pbio.1000178>.

O’Gorman, Eoin J., Owen L. Petchey, Katy J. Faulkner, Bruno Gallo, Timothy A. C. Gordon, Joana Neto-Cerejeira, J’on S. ’Olafsson, Doris E. Pichler, Murray S. A. Thompson, and Guy Woodward. 2019. “A Simple Model Predicts How Warming Simplifies Wild Food Webs.” *Nature Climate Change* 9 (8): 611–16. <https://doi.org/10.1038/s41558-019-0513-x>.

Patonai, Katalin, and Ferenc Jord’an. 2017. “Aggregation of Incomplete Food Web Data May Help to Suggest Sampling Strategies.” *Ecological Modelling* 352 (May): 77–89. <https://doi.org/10.1016/j.ecolmodel.2017.02.024>.

Peralta-Maraver, I., M. J. L’opez-Rodríguez, and J. M. Tierno de Figueroa. 2017. “Structure, Dynamics and Stability of a Mediterranean River Food Web.” *Marine and Freshwater Research* 68 (3). CSIRO PUBLISHING: 484–95. <https://doi.org/10.1071/MF15154>.

Petchey, Owen L, Andrew P Beckerman, Jens O Riede, and Philip H Warren. 2008. “Size, Foraging, and Food Web Structure.” *Proceedings of the National Academy of Sciences* 105 (11). National Acad Sciences: 4191–6.

Petchey, Owen L., Andrew P. Beckerman, Jens O. Riede, and Philip H. Warren. 2011. “Fit, Efficiency, and Biology: Some Thoughts on Judging Food Web Models.” *Journal of Theoretical Biology* 279 (1): 169–71. <https://doi.org/10.1016/j.jtbi.2011.03.019>.

Petchey, Owen L., P. Timon McPhearson, Timothy M. Casey, and Peter J. Morin. 1999. “Environmental Warming Alters Food-Web Structure and Ecosystem Function.” *Nature* 402 (6757): 69–72. <https://doi.org/10.1038/47023>.

Petchey, Owen L., Mikael Pontarp, Thomas M. Massie, Sonia K’efi, Arpat Ozgul, Maja Weilenmann, Gian Marco Palamara, et al. 2015. “The Ecological Forecast Horizon, and Examples of Its Uses and Determinants.” *Ecology Letters* 18 (7): 597–611. <https://doi.org/10.1111/ele.12443>.

Poisot, Timoth’ee, and Daniel B. Stouffer. 2016. “How Ecological Networks Evolve.” Preprint. Ecology. <https://doi.org/10.1101/071993>.

Rall, B. C., U. Brose, M. Hartvig, G. Kalinkat, F. Schwarzmuller, O. Vucic-Pestic, and O. L. Petchey. 2012. “Universal Temperature and Body-Mass Scaling of Feeding Rates.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1605): 2923–34. <https://doi.org/10.1098/rstb.2012.0242>.

Shriner, Daniel, Yi Liu, David C. Nickle, and James I. Mullins. 2006. “Evolution of Intrahost Hiv - 1 Genetic Diversity During Chronic Infection.” *Evolution* 60 (6): 1165–76. <https://doi.org/10.1111/j.0014-3820.2006.tb01195.x>.

Tamaddoni-Nezhad, Alireza, Ghazal Afroozi Milani, Alan Raybould, Stephen Muggleton, and David A. Bohan. 2013. “Construction and Validation of Food Webs Using Logic-Based Machine Learning and Text Mining.” In *Advances in Ecological Research*, 49:225–89. Elsevier. <https://doi.org/10.1016/B978-0-12-420002-9.00004-4>.

Toni, Tina, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P.H. Stumpf. 2009. “Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems.” *Journal of the Royal Society Interface* 6 (31): 187–202. <https://doi.org/10.1098/rsif.2008.0172>.

Tylianakis, Jason M., and Amrei Binzer. 2014. “Effects of Global Environmental Changes on ParasitoidHost Food Webs and Biological Control.” *Biological Control* 75 (August): 77–86. <https://doi.org/10.1016/j.biocontrol.2013.10.003>.

Warren, Philip H. 1989. “Spatial and Temporal Variation in the Structure of a Freshwater Food Web.” *Oikos* 55 (3): 299. <https://doi.org/10.2307/3565588>.

Williams, Richard J., and Neo D. Martinez. 2000. “Simple Rules Yield Complex Food Webs.” *Nature* 404 (6774, 6774). Nature Publishing Group: 180–83. <https://doi.org/10.1038/35004572>.

Woodward, Guy, Julia Blanchard, Rasmus B. Lauridsen, Francois K. Edwards, J. Iwan Jones, David Figueroa, Philip H. Warren, and Owen L. Petchey. 2010. “Chapter 6 - Individual-Based Food Webs: Species Identity, Body Size and Sampling Effects.” In *Advances in Ecological Research*, edited by Guy Woodward, 43:211–66. Integrative Ecology: From Molecules to Ecosystems. Academic Press. <https://doi.org/10.1016/B978-0-12-385005-8.00006-X>.