Inferring food web structure from predator–prey body size relationships

Dominique Gravel1,2*, Timothée Poisot1,2, Camille Albouy3,4, Laure Velez3 and David Mouillot3,5

1Département de biologie, chimie et géographie, Université du Québec à Rimouski, 300 Allée des Ursulines, G5L 3A1 Québec, Canada; 2Québec Centre for Biodiversity Science; 3UMR CNRS-UM2-IRD-IFREMER 5119 ECOSYM, Université Montpellier 2, CC 093, 34095 Montpellier Cedex 5, France; 4Laboratoire Ecosystèmes Marins Exploités UMR 212, IRD, IFREMER, UMI, UMI, avenue Jean Monnet BP171, 34203 Sete Cedex, France; and 5ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Qld, 4811, Australia

Summary

1. Current global changes make it important to be able to predict which interactions will occur in the emerging ecosystems. Most of the current methods to infer the existence of interactions between two species require a good knowledge of their behaviour or a direct observation of interactions. In this paper, we overcome these limitations by developing a method, inspired from the niche model of food web structure, using the statistical relationship between predator and prey body size to infer the matrix of potential interactions among a pool of species.

2. The novelty of our approach is to infer, for any species of a given species pool, the three species-specific parameters of the niche model. The method applies to both local and metaweb scales. It allows one to evaluate the feeding interactions of a new species entering the community.

3. We find that this method gives robust predictions of the structure of food webs and that its efficiency is increased when the strength of the body–size relationship between predators and preys increases.

4. We finally illustrate the potential of the method to infer the metaweb structure of pelagic fishes of the Mediterranean sea under different global change scenarios.

Key-words: Metaweb, Body size, Niche model, Food web

Introduction

One of the current challenges in ecology is to predict the emergence and re-assembly of communities following species responses to global changes. Understanding this is made all the more important as these novel ecosystems will be more common with increasing human pressure. We know that species invasions, biomass harvesting, ranges shifts, disturbances and changes in land use are important drivers of biodiversity turnover. How they affect species composition is now well described (Pereira et al. 2010), but forecasting their impact on community structure and functioning requires a priori knowledge of potential interactions among species. Predicting interactions among species that never co-occurred proves challenging, as traditional empirical methods of food web sampling such as stomach content analysis cannot be applied. While for competitive communities this task could be achieved using predictive modelling based on functional traits (McGill et al. 2006; Albouy et al. 2010) or phylogeny (Cavender-Bares et al. 2009; Mouquet et al. 2012), it is much more difficult to forecast trophic interactions in a food web (Tylianakis et al. 2008; Ings et al. 2009; Montoya & Raffaelli 2010).

Inferring potential interactions among species of an arbitrary defined pool is a major step to predict the structure of emergent communities and their functioning. We call these potential interactions among a given set of species, whether at the local or at the regional scales, the metaweb (Dunne 2006). A metaweb takes the form of an adjacency matrix $W$, of size $s \times s$ for a pool of $s$ species, and in which $W_{ij} = 1$ if species $i$ can consume species $j$, and 0 otherwise. This matrix aggregates the trophic interactions among all species from the pool that are susceptible to both co-occur and interact at the regional scale (Dunne 2006). With a metaweb in hands, one could analyse the impacts of global changes, such as range shifts or species invasions, on the potential community structure, by extracting the relevant species and examining the properties of the community they define. While this concept is progressively finding its way through theoretical spatial food web ecology (Pillai, Loreau & Gonzalez 2009; Lafferty & Dunne 2010; Gravel et al. 2011a, b), it is still limited by data availability and predictive accuracy. Published metaweb data rely on literature surveys (Havens 1992; Pechnik et al. 2008; Baiser et al. 2012), or a compilation of several local food webs (Stanko et al. 2002;
Poisot et al. 2012), making them resolved only for species that co-occurred a large enough number of times.

Development of predictive models of trophic interactions could greatly improve our understanding of large-scale food web structure and our capacity to anticipate major changes in ecosystem functioning. Theory should provide some guidance about the general rules underpinning interactions among a set of species and thus help us to infer the metaweb. The actual food web theory is largely derived from the niche model (Williams & Martinez 2000). This model simply and intuitively poses that each species in a food web has a niche position \( n_i \), a feeding niche optimum \( c_i \) and a range \( r_i \) of suitable preys around that optimum (Fig. 1). These simple rules are sufficient to generate realistic food web structures that fit most of published food webs (Dunne 2006). The niche model was a substantial improvement of the previous cascade model (Cohen, Briand & Newman 1990), and subsequent models (e.g. the nested hierarchy model (Cattin et al. 2004), the minimum potential model (Allesina, Alonso & Pascual 2008), the probabilistic niche model (Williams, Anandanadesan & Purves 2010) are somehow derived from these rules (Stouffer et al. 2005). Although it is based on different assumptions, the adaptive foraging theory of food web structure (Petcheys et al. 2008) provides comparable predictions (Williams et al. 2010). One major recent breakthrough in food web theory has been the attempt to parameterize the niche model and other food web models from field data and to compare their fit through likelihood methods (Allesina, Alonso & Pascual 2008; Williams et al. 2010; Williams 2011). These methods provide, for each species of the food web, the optimal parameters to fit the empirical web structures, given the hypothesized underlying rules of the niche model. Despite their theoretical interest, these methods, however, come with several drawbacks when comes the time to perform biodiversity scenarios. First, they are difficult to apply at large scale because of the technical and logistical requirements of metaweb data collection. Second, once the model is parameterized, it could only be used to infer feeding interactions between species with already documented interactions (i.e. it is impossible to infer potential interactions among species that do never co-occurred). Finally, the model optimization is a serious challenge for large data sets with a large number of parameters to evaluate and rough likelihood surfaces. There is consequently an urgent need for a method that could rapidly and easily provide an estimate of potential interactions in a metaweb based on incomplete data.

In this paper, we present a method designed to infer the potential interactions between all pairs of species from a given pool based on observations of body size of predators and their prey. The method applies to both the local and the metaweb scales. We do so by a parameterization of the niche model, based on the well-documented allometric scaling relationship between predator and prey body size (Cohen, Jonsson & Carpenter 2003; Brose, Williams & Martinez 2006; Riede et al. 2010). We first develop the method and apply it to food webs from various environments. We find that the method accurately predicts interactions (and lack thereof) and that its accuracy increases with the strength of the predator–prey body size relationship. We then analyse the sensitivity of the method to incomplete data (missing links) and find that it is robust to sampling effort. We finally illustrate the potential of the method to infer the metaweb structure of pelagic fishes of the Mediterranean sea and the consequences of alteration of body size distribution by global changes or anthropic forcings. The method is best suited for strongly size-structured food webs and will likely not hold for other types of non-body-size-structured interactions. We therefore conclude on the future issues to generalize this approach to other types of trait-matching models.

**Model description**

The method aims to infer the potential interactions among a pool of species from a subset of observations of predator–prey interactions. The method follows the following steps, with details provided below:

**Step 1**: Log transformation of the body size data;

**Step 2**: Statistical analysis of the predator–prey body size relationship;

**Step 3**: Inference of the niche model parameters for all species from the species pool;

**Step 4**: Interpretation of the parameters and computation of the metaweb.
We also provide an example of R code (R Core Development Team) and data in the Supplementary Material, detailing the step by step procedure and the format of the input data.

INFERRING PARAMETERS FROM THE NICHE MODEL AND BUILDING THE METAWEB

The niche model predicts the food web structure from a set of three species-specific parameters (Williams & Martinez 2000): the niche position \( n_i \), the feeding niche optimum \( c_i \) (called the centroid) and the feeding range \( r_i \). A species \( i \) will predate all species \( j \) whose niche position \( n_j \) lies within the interval \([c_i - r_i/2, c_i + r_i/2]\) (Fig. 1). We evaluate all of these parameters from the predator–prey body size relationship, enabling us to parameterize the metaweb from observed interactions only. The parameterization is robust to the sampling effort, as it is much easier to document interactions than their absence (Martinez et al. 1999). Our approach is, however, mostly limited to predatory interactions since the body size relationship between herbivores and primary producers is not as general (Riede et al. 2010) and obviously do not hold for parasitic, mutualistic or competitive networks.

Assuming that body size is the main niche axis structuring trophic interactions, the parameter \( n_i \) corresponds simply to the log of body size (in mass or length) of species \( i \). Though only the relative position of all species along the body size gradient needs to be respected, it is possible to standardize log body size values between 0 (minimum size in the regional species pool) and 1 (maximal size). We then consider a linear relationship between the decimal logarithm of body size and the centroid of the niche (the dark line at Fig. 1). This relationship is obtained by fitting the linear model

\[
c = \log_{10}(M_{\text{prev}}) = a_0 + a_1 \times \log_{10}(M_{\text{pred}})
\]

to the data, where \( M_{\text{prev}} \) and \( M_{\text{pred}} \) are the prey and predator body size, respectively. The lower and upper boundaries of the feeding range are easily obtained by fitting the 5% and 95% quantile regressions between \( \log_{10}(M_{\text{prev}}) \) and \( \log_{10}(M_{\text{pred}}) \) (the dotted lines at Fig. 1, see the example at Fig. 2). We note them as \( r_{\text{low},i} \) and \( r_{\text{high},i} \) respectively, and the corresponding parameters of the linear quantile regressions are \( \beta_{0,\text{low}}/\text{high} \) and \( \beta_{1,\text{low}}/\text{high} \). In sum, the parameter \( n_i \) for any species of the metaweb is given by the standardized value of the log body size \( M_i \), \( c_i \) is estimated from the linear regression between predator and prey log body size and \( r_i \) from the quantile regressions. Type I regressions were used because there is no equivalence of type II models for quantile regression.

The next step of the methodology is to reconstruct the metaweb. Once these parameters are calculated from a subsample of species from the regional pool, the coefficients of the different linear models are used to infer the niche parameters of each species of the species pool. Again, the niche parameter for any species \( i \), \( n_i \), is given by the log of body size. The centroid of the niche is obtained by the relationship \( c_i = a_0 + a_1 \times n_i \), the lower boundary of the niche is \( r_{\text{low},i} = \beta_{0,\text{low}} + a_1,\text{low} \times n_i \) and the upper boundary \( r_{\text{high},i} = \beta_{0,\text{high}} + a_1,\text{high} \times n_i \). A feeding link from species \( i \) to species \( j \) occurs if \( n_j > r_{\text{low},i} \) & \( n_j < r_{\text{high},i} \). We provide an example at Fig. 2.

**Method accuracy**

**PREDICTIVE PERFORMANCE**

We illustrate the method with the food web data sets of Brose et al. (2005). The meta-analysis of Brose et al. (2006) was conducted on this data set to test the generality of the predator–prey body size relationship across different systems (terrestrial, aquatic and marine). The relationship was found to be strong across all systems despite exhibiting variability from one to another. Each web has between 26 and 380 species and 18 and 1466 feeding links. Several of these webs are repetitions over time at a single location, in which cases we pooled the data for each of the 15 different locations to calculate the predator–prey body size relationship. We removed 4 data sets that had a non-significant predator–prey relationship and were thus useless with our approach. The links are not systematically sampled, meaning that any absence of a link between two species for a given data set could either be a real absence or due to insufficient sampling or lack of information. While the predator–prey body size relationship is strong over all data sets (Brose, Williams & Martinez 2006), there is quite substantial variation among them, enabling us to assess the sensitivity of the method to the strength of this relationship.

We assessed the performance of our method using the True Skill Statistic (TSS). The TSS is based on the partition of events (the prediction of a trophic interaction) between four components: the component \( a \) reports the number of links that are both predicted and observed, \( b \) reports predicted links with no corresponding observation, \( c \) reports the number of observed links that are predicted absent and \( d \) reports the number of predicted and observed absent links. The TSS is then
calculated as $TSS = \frac{(ad-bc)}{((a+c)(b+d))}$. The $TSS$ quantifies the proportion of prediction success relative to false predictions and returns values ranging between 1 (perfect predictions) and $-1$ (inverted forecast) (Allouche, Tsoar & Kadmon 2006).

We calculated the $TSS$ for each of the 11 different webs and related it to the strength of the predator–prey body size relationship, measured by the $R^2$ of the linear model. We find that the $TSS$ is positive for all webs, ranging from 0·13 to 0·76 (Fig. 3a). We find a positive relationship between the $R^2$ of the linear model and the $TSS$ ($R^2 = 0·50, P = 0·016$). When we decompose the different components of predictions and observations, we find that the fraction of prediction match is high, with an average of $(\bar{a} + \bar{d})/S^2 = 0·58$, Fig. 3b). The fraction of wrong predictions is lower, at $\bar{b} + \bar{c}/S^2 = 0·40$ and decreases with the $R^2$ of the predator–prey body size relationship.

The parameterized niche model tends to overestimate the number of links in a web (see the example at Fig. 2). This result is significant but not surprising, given that these data sets do not necessarily contain all links, as they were not designed with this purpose, and thus, some of the links might have been missed. This interpretation is also reminiscent of previous debates on the difficulty to sample all links in a web (Martinez et al. 1999). It is also well known that the niche model predicts a continuous diet along the niche axis (the webs are said to be
interval (Cohen, Briand & Newman 1990; Stouffer, Camacho & Amaral 2006), while real food webs do not have this characteristic. We thus might over predict link density within the niche of a given species. Previous studies (Allesina, Alonso & Pascual 2008) and the Application 1, however, show that this problem is easily circumvented when a second niche axis, e.g. an environmental niche, comes into play. A recent study on dimensionality of networks shown that most webs have between 3 and 6 dimensions (Eklof et al. 2013), meaning that adding a few more niche axes can greatly improve the accuracy of predictions.

SENSITIVITY TO SAMPLING EFFORT

We subsequently explored the impact of sampling effort on the accuracy of the model. We assumed a random subsampling of all the interactions occurring in the food web. To do so, we selected a species-rich food web ($S = 67$) from the Brose et al. data set, with 601 observed feeding links and a good TSS (0.51). We randomly removed from 0 to 90% of the observed links to do the evaluation of the parameters of the linear regressions (i.e. decrease the quantity of information used to calibrate the model). We then after compared the empirical web (with all species) with the inferred web with these parameters (all species also). The comparison was done again with the TSS. We performed 100 randomizations per number of removed links. This numerical simulation reproduces incomplete sampling in the process of building the food web. We find that the TSS remains constant up until 80% of the observed links are removed (Fig. 4). At this level, the TSS starts to decline drastically and its variance increases.

This result shows that our method of parameterization of the niche model is robust with regard to sampling effort. Note, however, that a biased sampling with respect to body size (e.g. sampling of the largest species) might be more likely to reduce the fit of the predator–prey body size relationship. The aggregation intro trophic species that do have the same body size and the same diet should not impact the prediction since it will not affect parameter evaluation. It will, however, reduce the fit of the predator–prey body size relationship and therefore the accuracy of the method if the trophospecies do have different body sizes. The same artefact will be observed if there is strong intra-specific variation in body size (see the Discussion).

Application: Mediterranean food web structure under fishing pressure

DATA SET

We now present an application of the method to infer the metaweb of interactions among fishes of the Mediterranean sea. The Mediterranean is known as a hotspot of fish diversity that is severely threatened by climate change and overfishing (Mouillot et al. 2011). There are 557 fish species in the regional pool, with a maximum body size ranging from 2.3 to 1100 cm (Whithead et al. 1986; Louisy 2005; Froese & Pauly 2011). Chondrichthians, mammals and turtles were under represented in the two trophic networks and were removed from the analyses. We parameterized the niche model with a subset of species, from two different highly resolved food webs from the Catalan area (Coll et al. 2006) (82 species) and Corsica (Albouy et al. 2010) (58 species).

INFERRING THE METAWEB FOR MEDITERRANEAN FISHES

We estimated parameters $n$, $c$ and $r$ for each of the 557 species and inferred the potential interactions among all of them (the example in the Supplementary code is based on this data set). The metaweb has a total of 126 501 links, for a connectance of 0.41 (Fig. 5a). The metaweb is also highly nested (specialist species feed on a subset of prey of the most generalist species. Fig. 5a). We also considered a second niche axis related to species spatial distribution. Most fish species have restricted geographic range within the Mediterranean sea because of specific response to temperature and other environmental variables (Albouy et al. 2012). We therefore removed from the metaweb all links between species having no range overlap. Data on the extent of occurrence of fish species were compiled from a published atlas of fishes of the northern Atlantic and the Mediterranean (Whitehead & Unesco 1986). This atlas is based on regional data sets and expert knowledge and was edited between 1984 and 1986. It currently provides the only available basin-wide information on the extent of occurrence of all Mediterranean Sea fish species. The above-mentioned atlas do not account for the bathymetric distribution of Mediterranean fish species, yet bathymetry is
considered as one of the main factors accounting for marine fish distributions (Louisy 2005). We therefore refined the extent of occurrence maps by clipping off areas with depths that fall outside the minimum or maximum known for the species. Species’ bathymetric ranges were obtained from FishBase (Louisy 2005; Froese & Pauly 2011). The resulting metaweb has a total of 95 989 links, for a connectance of $C = 0.31$. Connectance decreases because links are removed by incompatibilities in bathymetry but the total number of species stays constant. This metaweb is clearly less interval (Fig. 5b). Contiguous gaps in the diet are likely to emerge from modularity in the co-occurrence matrix (Araújo et al. 2011). Future studies should explore how co-occurrence is constraining the topological structure of metawebs.

**IMPACT OF GLOBAL CHANGES**

Our method to parameterize the niche model has a unique feature relative to the original niche model (Williams & Martínez 2000): the network properties of the original model, such as the number of links, food chain length and degree distribution (the number of in and out feeding links per species), are controlled by the input parameters (species richness and connectance), where all properties of our parameterized niche model are emergent features of the predator–prey body size relationship and the frequency distribution of body size. The connectance in the niche model is fixed by the user; in our case, it strongly depends on the scaling of the feeding range with body size and the frequency of larger bodied species.

This feature is particularly important to understand global change impacts on community structure. For instance, a common prediction of the impact of fisheries on the body size distribution is the reduction of the average and variance of body size (Jackson et al. 2001). Similarly, global warming is also expected to alter fish body size distribution towards smaller species (Cheung et al. 2012). We explored by simulations the impact of these changes on the degree distribution of the Mediterranean metaweb. Results are illustrated at Fig. 6. The shape of the cumulative degree distribution provides a visual assessment of the distribution of diet specificity in a network. The curves becomes steeper and the fall moves to the left with increasing specialization (indicative of a larger proportion of species with a low degree). Even if the simulated scenarios are crude representations of the predicted alterations of community structure, they both show that fisheries impacts on body size will substantially alter the network properties. The results at Fig. 6 show that reduction in the average body size will decrease the average number of preys per predator (there is an initial sharper decline of the cumulative distribution), but that super generalist species will also appear (because of the shift of the tail towards the right). The removal of the 40% largest species reduces much more the average generality which is expected as larger species are predators with a large feeding niche in the niche model. The two scenarios consequently increases substantially the relative abundance of species with a smaller degree, even for the average size reduction scenario where the total number of species is held constant. The change in the degree distribution is likely to reduce substantially the expected persistence and stability of these communities (Gravel et al. 2011a,b).

**Discussion**

In this paper, we presented a method to infer potential interactions among an arbitrary pool of species. The data required to perform this methodology are simple to obtain, as the body size of a large number of species are available from reference databases or easy to measure. Because we rely on a robust allometric relationship, applying this method requires neither complex statistical techniques nor an exhaustive knowledge of the realized interactions within the metaweb of interest. In addition to the opportunity to simulate biodiversity scenarios, our method can also help generating baseline expectations about the food web structure for environments which are notoriously difficult to sample, such as soils, deep-sea environments or fossil records. However, because this allometric relationship is characteristic of predatory interactions, our method will likely not hold for other types of non-body-size-structured interactions such as herbivory and parasitism.

It appears in Fig. 2 that the feeding range increases with the body size of the predator. This result was quite general across data sets with a good fit of the predator–prey body size relationship, although with considerable variation in the slope of the relationship. It offers the possibility to compare empirical-based estimates of the parameters to the original niche model of Williams & Martínez (2000). For this specific data set, the model for the centroid is $c_i = -1.34 + 0.66n_i$ and the average range ($\bar{r}_i = (r_{\text{high},i} - r_{\text{low},i})/2$) is $\bar{r}_i = 0.63 + 0.78n_i$. The original niche models assumed that $r_i$ scales linearly with connectance $C$ and the niche position, such that $r_i = 2Cn_i$. Even though the parameterization suggests a linear increase in generality with body size, the final number of preys in the diet for a given size will depend on the frequency distribution of body size. A large predator for instance could have a large feeding range, but only few species to find on and effectively appears as a specialist. This relationship will have to be investigated further, among several food webs (e.g. Digel et al. (2011)), because the scaling of generality with trophic rank has several consequences on persistence (Gravel et al. 2011a) and food web dynamics (Brose et al. 2006; Berlow et al. 2009).

Ontogenic shifts in diet are common in many size-structured populations and could be accounted for in two ways with this methodology. First, they will enter the model parameterization by simply including a link from species A (the largest) to species B (the smallest) and the opposite. This would yield in a data point for species B figuring above the 1:1 relationship. Such a data point will obviously influence the model calibration and enlarge the regime of all species. It is not obvious, however, that ontogenic shifts will be found in the metaweb as it would require an upper limit of the range above the 1:1 relationship. The alternative approach would be to distinguish ‘subspecies’ by size categories in the model calibration and then in the reconstruction of the metaweb. This approach would force
ontogenic shifts and perhaps more precisely represent the interaction matrix where the within-population size structure is important.

The method should be completed with other sources of information to better predict gaps in the interaction matrix. The original niche model was definitely inspired by the predator–prey body size relationship, but was intended to be more general and eventually deal with several niche axes (Williams & Martínez 2000). Allesina, Alonso & Pascual (2008) indeed found that adding a second axis, creating gaps into the first axis & Martinez 2000). Allesina, Alonso & Pascual (2008) indeed issue because there is an urgent need to include interactions succession of species diets along the niche axis), a structural issue general and eventually deal with several niche axes (Williams variation we found in our analysis, for instance by constrain-

tor that traits are conserved along the phylogeny (Mouquet
cal efforts will also have to take into account more various and heterogeneous sources of data such as phylogeny (Eklof
equality to constrain the prior distribution of the likelihood of any pairwise interaction, based on trait matching, using alternative source of data. The method we employed here was based on the match between predator and prey traits. This method could be generalized by looking at the match between all potential traits. Eklof et al. (2013) indeed found that the match between only three traits is enough to encompass a vast majority of interactions. But when such data is not always available, then it might be possible to use phylogeny as a proxy for a high-dimensional trait species, provided that traits are conserved along the phylogeny (Mouquet et al. 2012). A bayesian addition to the model would be to weight the likelihood that species \( i \) is a predator of species \( j \), considering that the closest relative of this predator is also a predator of \( i \). It could account for part of the unexplained variation we found in our analysis, for instance by constraining the range of the diet for some large predators or by creating holes within the feeding range (Rall et al. 2011).

Adding this method to the ecologist’s toolbox is a timely issue because there is an urgent need to include interactions into future biodiversity scenarios (Thuiller et al. 2013). There has been impressive progress in biogeography over the last fifteen years to predict the impact of global changes on range shifts, extinction risks and the future distribution of biodiversity, but we now face the realization that emerging ecosystems are unlikely to be a spatial displacement of the ones we currently know. This method will help assessing global change impacts of community structure, moving the field of biogeography closer to ecosystem functioning.

Acknowledgements

We thank I. Gournard for preparing the food web data. We also thank N. Martínez, U. Brose and two anonymous reviewers for their constructive comments and stimulating discussion. DG received financial support from NSERC and the Canada Research Chair program. TP and CA are funded by a MELS-FQRNT postdoctoral fellowship.

References


Louisy, P (2005)


Ings, T.C., Montoya, J.M., Bascompte, J., Bluthgen, N., Brown, L. & Dormann,


Received 27 March 2013; accepted 15 July 2013
Handling Editor: Robert Freckleton

Supporting Information
Additional Supporting Information and R code are available online on FigShare at DOI: http://dx.doi.org/10.6084/m9.figshare.650228