Effects of data composition and diversity on precision and accuracy of network inference

11 May 2021

Contents

1 Abstract

2 Introduction

2.1 Why is knowing trophic interactions important?

Knowledge about the trophic interactions in a food web can be used to answer crucial ecological questions: Which species is a keystone species in a food web (Jordán 2009)? How would a food web respond if the top predators are removed (Knight et al. 2005)? Which species should be given priority when conservation efforts are implemented (McDonald-Madden et al. 2016)? How robust is a food web to extinctions (Dunne, Williams, and Martinez 2002)?

2.2 What are the methods for knowing interactions?

Species interactions can be inferred from different types of methods (Table 1) such as gut content analysis (Peralta-Maraver, Lopez-Rodriguez, and de Figueroa 2017), stable isotope ratio analysis of tissues (Layman et al. 2007), experimentation (Warren 1989) and DNA metabarcoding (Roslin and Majaneva 2016). The types of information extracted from these methods vary. E.g. Gut content analysis of a predator results in the identification of prey species fed upon by that predator, while the stable isotope ratio analysis of the organism's tissue can help us infer the approximate trophic position of that species in the food web.

2.3 Gut content analysis

Gut content analysis is performed to know what an individual has consumed recently. The diet of that individual can be inferred from the items in the stomach. There is considerable uncertainty involved in this method (Baker, Buckland, and Sheaves 2014). E.g. In fish, there is ambiguity while separating prey items in the gut. There are loose tissues that are not identifiable and cannot be assigned to a specific prey item with certainty. There are factors such as sample size of consumers, mechanical prey handling, differential digestion and evacuation rates of different types of different prey types and volumes, and the ingestion order that in combination result in an unquantifiable error which is difficult to interpret in the predator diet (Hyslop 1980; Rindorf and Lewy 2004; Baker, Buckland, and Sheaves 2014).

2.4 Stable isotope analysis

Stable isotope composition of a specific element in a consumer's tissue is given by the ratio of a rarer, heavier isotope to a more common lighter isotope (McCormack et al. 2019). Stable isotope analysis (SIA) of tissues from a single individual can be used to quantify its diet over varying temporal and spatial scales from one sample (Crawford, Mcdonald, and Bearhop 2008). Using SIA, several different sources of information can be extracted from a single sampling event as different tissues integrate diet over different temporal scales as well as spatial scales. However, there is uncertainty involved in this method which is due to 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).

Trophic level of a species $i(TL_i)$ is calculated as:

$TL_i = \frac{\delta^{15}N_i - \delta^{15}N_{base}}{\Delta^{15}N} + TL_{base}$

where $\delta^{15}N_i$ is the $\delta^{15}N$ value of species i, and TL_{base} and $\delta^{15}N_{base}$ are known TL and $\delta^{15}N$ values corresponding to a reference baseline species from a given food web. $\Delta^{15}N$ is the trophic fractionation value (usually 3.4%)

2.5 Why inferring interactions is not easy?

As a food web can be quite complex with many species and trophic interactions (Williams and Martinez 2000; Berlow et al. 2009), inferring these trophic interactions with high accuracy and low uncertainty would require large amount of food web data (citreqd). E.g. One would need to sample many individuals for gut

Table 1: Studies showing the type of data used to infer trophic interactions in a food web

Data type	Study	Type of information inferred
Gut contents	O'Malley et al. (2017)	Predator diet
	Dixon et al. (2017)	Predator diet
	McClain-Counts et al. (2017)	Predator diet
	Peralta-Maraver et al. (2016)	Predator diet
Stable isotope composition		

content data, and processing these guts require a large amount of effort and time as well. Hence, there needs to be prioritisation of sampling based on an individual trait which is easily identifiable in the field, which would lead to high accuracy in predicting food web with minimal effort.

Even though when one had finally made a decision to use a single type of food web data to infer the species interactions, it might not be practically feasible to infer the complete food web structure from only a single type of food web data. E.g. Stable isotope analysis from tissues can only provide the approximate trophic position of the species which might not be sufficient to predict the complete food web structure. Also, gut content data might only be used to infer partial diet of a predator individual because gut content data only provides the information of the prey individuals that have been foraged by the predator recently. There is also a possibility that some of the information provided by these types of food web data are redundant i.e. taking into account any further food web data does not necessarily increase the food web prediction significantly. E.g. This can happen if the guts of two predator individuals give same information about the prey the predator has consumed.

Gut content data and stable isotope data provide different types of information about a food web on different taxonomic, spatial and temporal scales which results in different amount of uncertainties in inferring the species interactions. E.g. Gut content data provides information about the diet of an individual on a local scale whereas stable isotope provides the information about the diet on a bigger spatial scale which is dependent upon the amount of space dispersed by that species. Therefore, it is crucial to integrate these diverse information simultaneously from these methods so that information from different taxonomic, spatial and temporal scale can be integrated, requiring less sampling effort and resulting in a high accuracy of the food webs.

2.6 What we do in our study and how?

These questions are addressed *in silico* by simulating food web networks whose true parameters are known. We used the allometric diet breadth model (ADBM) to simulate the food webs and approximate Bayesian computation (ABC) to parameterise the ADBM. We simulated the food web data which were used to parameterise the ADBM using ABC. We also used real gut content data from Broadstone Stream food web to address those questions.

3 Materials and Methods

Firstly, we present an *in silico* approach (Fig. 1) to address the research questions where we simulated food web structure using the allometric diet breadth model (ADBM) whose true parameters were known. In the upcoming sections, we present methods to simulate gut content data and stable isotope ratio data with uncertainty. We vary the amount of simulated data and use partial data to predict food web using the ADBM parameterised via the rejection approximate Bayesian computation (ABC). Secondly, we used a gut content data from an observed Broadstone Stream food web to infer the trophic interactions.

3.1 Gut content data simulation

We simulated a food web using the ADBM for a given set of parameters. For a given set of predators, we subset the diet from the simulated food web. Then using a probability mass function (distribution), we
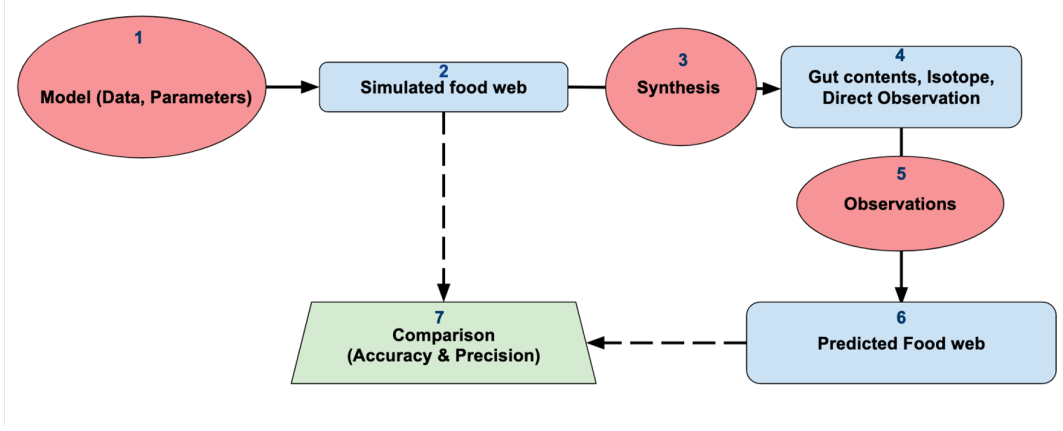
Figure 1: *In silico* approach for investigating the effect of quantity and composition of food web data in predicting food web structure using the ADBM.

sampled the gut content data from predators' diet thereby incorporating the uncertainty in the gut content data. We repeated this process multiple number of times for every predator in the food web.

Input:

- Predators whose diet are to be simulated $P = \{p_1, \ldots, p_n\}$

- A simulated food web $ADBM(\theta_i) = \{d_{p_1}, d_{p_1}, \ldots, d_{p_k}\}$, where d_{p_k} is a one-dimensional diet matrix of predator k containing ones and zeros.

- A function which describes uncertainty in the diet $U(d)$

- Number of independent guts to be simulated for a predator $p_i : ngut$

Sampling:

- for $p_i \in P$

  − for $j = 1, \ldots, ngut$, where $ngut$ is the number of guts to be simulated

    ∗ Simulate a single gut of a predator $p_i : g(p_i) = d_{p_i} * U(d_{p_i})$

  − Set of gut of a predator $p_i : G(p_i) = \{g(p_i) : g(p_i) = d_{p_i} * U(d_{p_i})\}$

Output:

- We simulated a pool of gut content data which contains simulated gut content data $G(p_i)$ for every predator p_i

3.2   Inferring food web using gut content data

From a pool of gut content data, we draw a set of gut content data randomly. Then we used the rejection ABC to accept a parameter value which would have resulted in the minimum distance, where distance = 1 - True skill statistic and the true skill statistic was computed between the predators' diet predicted from the ADBM and the sampled gut content data. We repeated this process for n ($= 100$) number of times for every i number of guts, where i lies between 1 and total number of gut content data in the pool.

Input:

- Predators $P : P = \{p_1, p_2, \ldots, p_k\}$

- A pool of gut content data $G : G = \{g_1, g_2, \ldots, g_n\}$, where g_n is a one-dimensional matrix containing ones and zeros.

4

- A model prediction $model(\theta) : ADBM(\theta) = \{d_{p_1}, d_{p_1}, \ldots, d_{p_k}\}$, where d_{p_k} is a one-dimensional diet matrix of predator k containing ones and zeros.

- A summary statistic $s(x) : s(x) \subseteq model(\theta)$

- A distance function $d(x_i, y) : d(x_i, y) = 1 - TSS(x_i, y)$

- An observed food web $Y : Y = \{d'_{p_1}, d'_{p_1}, \ldots, d'_{p_k}\}$

Sampling:

for $i = 1, \ldots, tgut$ where $tgut$ is the total number of gut content data in the pool G

- for $j = 1, \ldots, nsample$ where $nsample$ is the number of independent samples to be drawn

    - Draw a set of gut content data $y = \{g_1, g_2, \ldots, g_i\}$ from the pool of gut content data G
    - for $k = 1, \ldots, npar$ where $npar$ is the number of parameter values to be sampled
        * Draw a set of parameter values θ_k from the prior distribution $\pi(\theta)$
        * Compute the model result $x_i = model(\theta_k)$
        * Compute $s(x_i)$ and $d(s(x_i), y)$
    - Accept θ_j, which results in the $min_i\{d(s(x_i), y)\}$

- Compute $TSS_i(x, y) = \{TSS(x_i, y) : x_i = ADBM(\theta_j), \theta_j$ computed from previous step$\}$ using the accepted $\theta_1, \ldots, \theta_{nsample}$

Output:

A prediction interval containing TSS between observed and predicted food webs for every i number of gut content data drawn from the pool of gut content data.

3.3 Broadstone Stream food web

Broadstone Stream (51° 05' N 0° 03' E; 120 m above sea-level) is a second-order tributary of the River Medway in south-east England. It is dominated by invertebrates as the acidity of the stream (pH 4.7-6.6) excludes fish. The stream consists of 25 common invertebrate species (Woodward, Speirs, and Hildrew 2005a). Among the predators, the three large species are *Cordulegaster boltonii* Donovan, *Sialis fuliginosa* Pict. and *Plectrocnemia conspersa* [Curtis] and the three small species are the larvae of the tanypod midges *Macropelopia nebulosa* [Meigen], *Trissopelopia longimana* [Staeger], and *Zavrelimyia barbatipes* [Kieffer]. Broadstone Stream food web is one of the most completely described food webs for any system (Schmid-Araya et al. 2002; Hildrew 2009; Layer et al. 2010; and Woodward, Speirs, and Hildrew 2005b).

The Broadstone Stream food web in our study was constructed by aggregating by size (Woodward et al. 2010a) i.e a node is defined as collection of individuals within a specified size range. There were a total of 1008 gut content data. The number of prey items in predator's gut varied within a predator size class (Fig. 2). A single gut cannot give information about the complete diet of a predator as can be observed from the right skewed distribution (Fig. 2).

3.4 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). The ADBM predicts the set of prey species a consumer should feed upon to maximise its rate of energy intake (Petchey et al. 2008). The foraging variables in the model: energy content of the resources, handling times of the prey, space clearance rate and prey densities are allometrically scaled to the body sizes of the species.

The model correctly predicted a greater proportion of links when the trophic interactions were more strongly dependent on size (Petchey et al. 2008). Indeed, constructing a food web based only on body size (i.e. ignoring taxonomy) resulted in almost twice the number of correctly predicted links, i.e. 83%, in contrast to taxonomy (Woodward et al. 2010b). The goodness of fit of the ADBM's predictions depends on the types of interactions in the food webs (Petchey et al. 2008), because some of the interactions are more size structured than other interactions. For instance, predacious and aquatic herbivore interactions were predicted better than parasitoid and herbivory ones (Petchey et al. 2008).
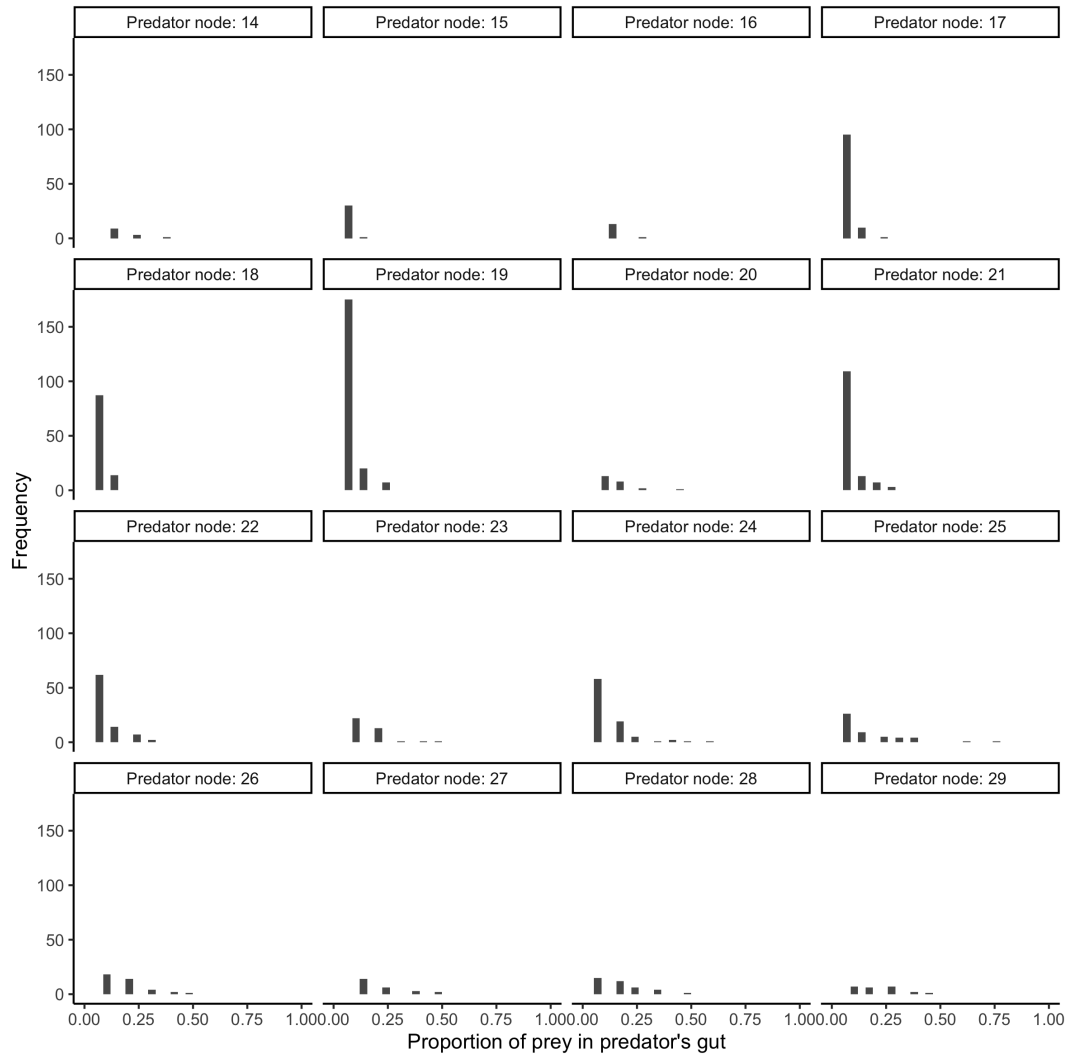
Figure 2: Distribution of proportion of prey in the gut of a predator belonging to a specific size class in the Broadstone Stream food web. The proportion of prey in a predator's gut is calculated as number of unique prey items observed in the predator gut divided by the total number of unique prey items in all the gut of that predator.

3.5 Assessment of prediction

The accuracy of the predicted diet of the predators was defined using true skill statistic (TSS) which takes into account the true and false predictions of both the presence and absence of links defined as:

$$\text{TSS} = \frac{ad - bc}{(a+c)(b+d)}$$

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

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

4 Results

We first present the ...

4.1 Prediction using real gut content data from the Broadstone Stream food web
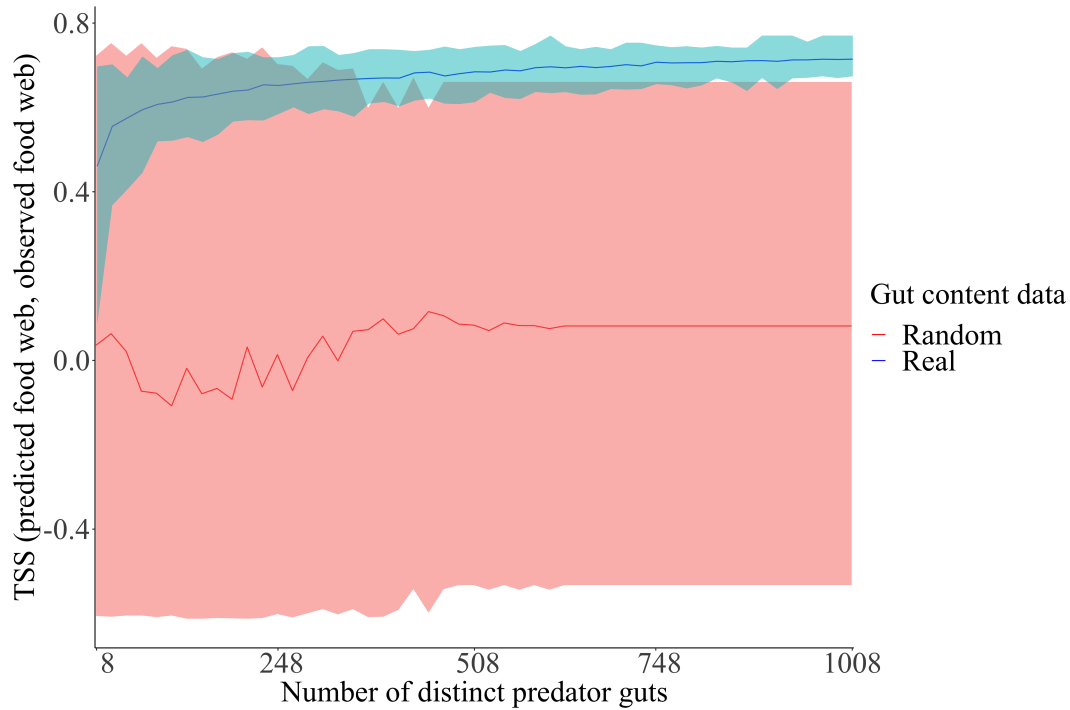

Figure 3: True skill statistics between the predicted and observed Broadstone Stream food web computed for number of different predator guts sampled. There were a total of 1008 guts from 29 species. Line represents the mean and shaded region represents the range/uncertainty. Blue and red curve/region correspond to real and randomly generated gut content data respectively.

The value of true skill statistics between the predicted food web and observed food web saturated quickly with the increasing number of distinct real predator guts (Fig. 3). With only 228 real predator guts, the ADBM parameterised with ABC predicted the food web with an accuracy of 90% of the maximum mean TSS (0.71). The uncertainty (red shaded region) in the TSS reduced with the increasing number of predators'

7

guts. Although at a low number of predators' guts the maximum TSS was quite high the uncertainty was high as well. This was quite possibly just by chance which was suggested by the overlap of the TSS region with that of randomly generated gut content data.

True skill statistics between the predicted food web and the observed food web had a saturating curve with the increasing number of distinct predator guts (Fig. 4) for both sets of gut predator guts: First set of gut data consisted of predator body sizes which were smaller than the mean body size and second set of gut data consisted of predator body sizes which were larger than the mean body size. However, the mean TSS corresponding to the guts of predators smaller than the mean body size was higher when compared to that from the guts of predators larger than the mean body size.
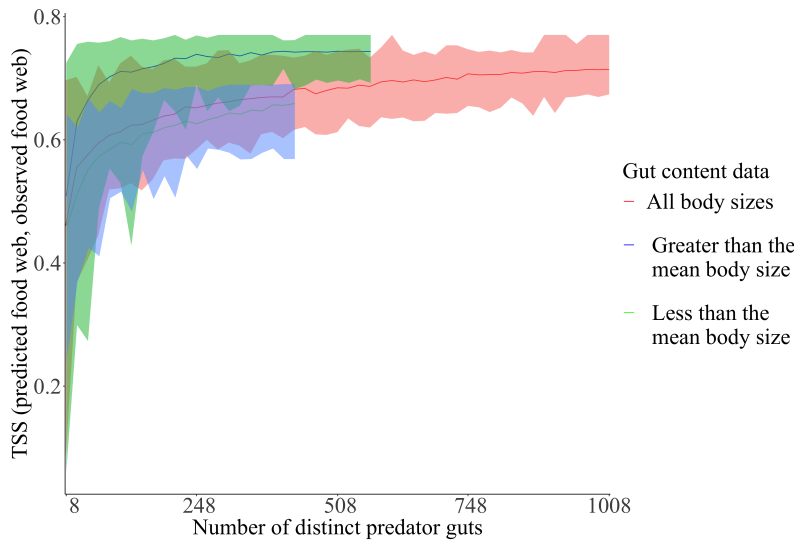

Figure 4: True skill statistics between predicted food web and observed food web estimated for different number of distinct predator guts. The estimation is done for three sets of gut data: gut content data of predators whose body sizes are smaller than the mean body size, larger than the mean body size, and all the gut content data.

4.2 Accuracy of model against accuracy of data

4.3 Mean standardised error in food web properties

4.4 Prediction using simulated stable isotope ratio data from a simulated food web

4.5 Prediction using both simulated stable isotope ratio and simulated gut content data from a simulated food web

5 Discussion

5.1 Why did we achieve a higher TSS when we use gut content data for predators less than mean body size?

The predator diets in the observed food web is not nested i.e. the diet of smaller species is not subset of the diet of larger species. When diets of predators with body size lower than mean body size is used to infer the food web, the predicted food web contains the diets of bigger predators as well because the
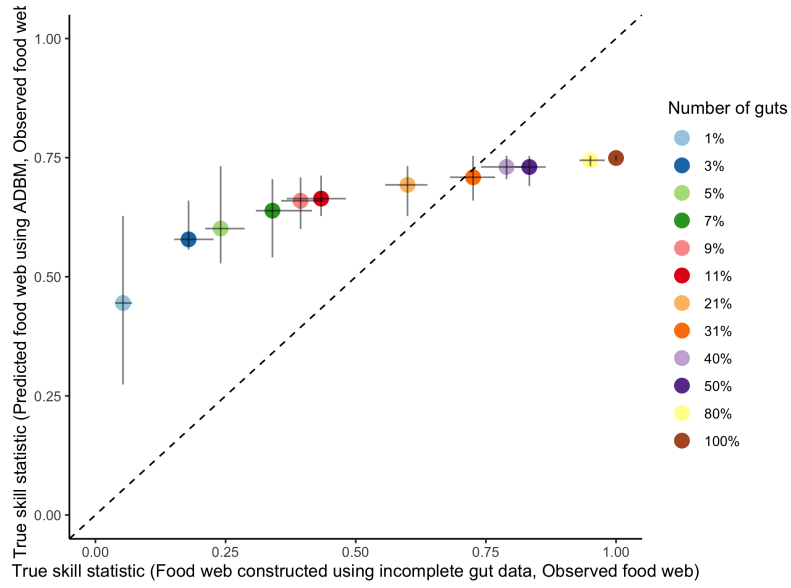
Figure 5: True skill statistic between predicted food web using ADBM, and observed food web plotted against the true skill statistic between food web constructed using incomplete gut content data, and observed food web.
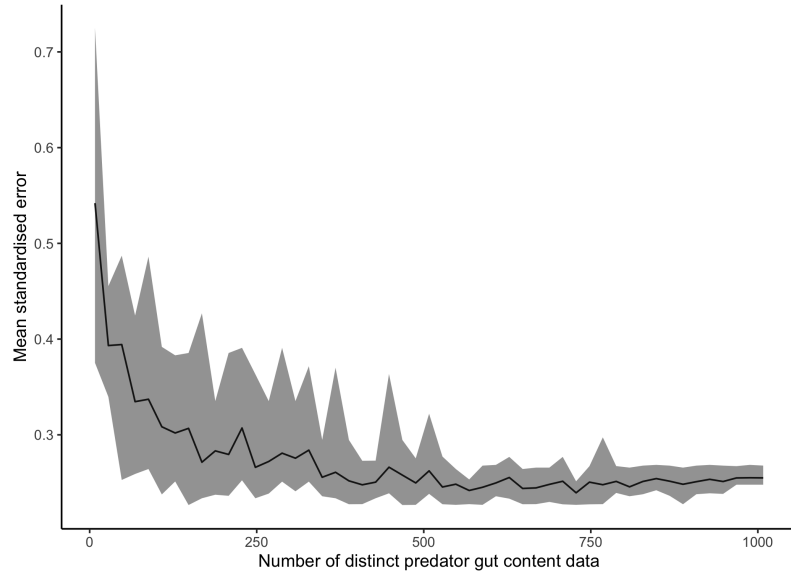

Figure 6: Mean standardised error for structural food web properties plotted against the number of distinct predator gut content data

5.2 Extension to other food web models

6 Conclusion

7 Supplementary Information

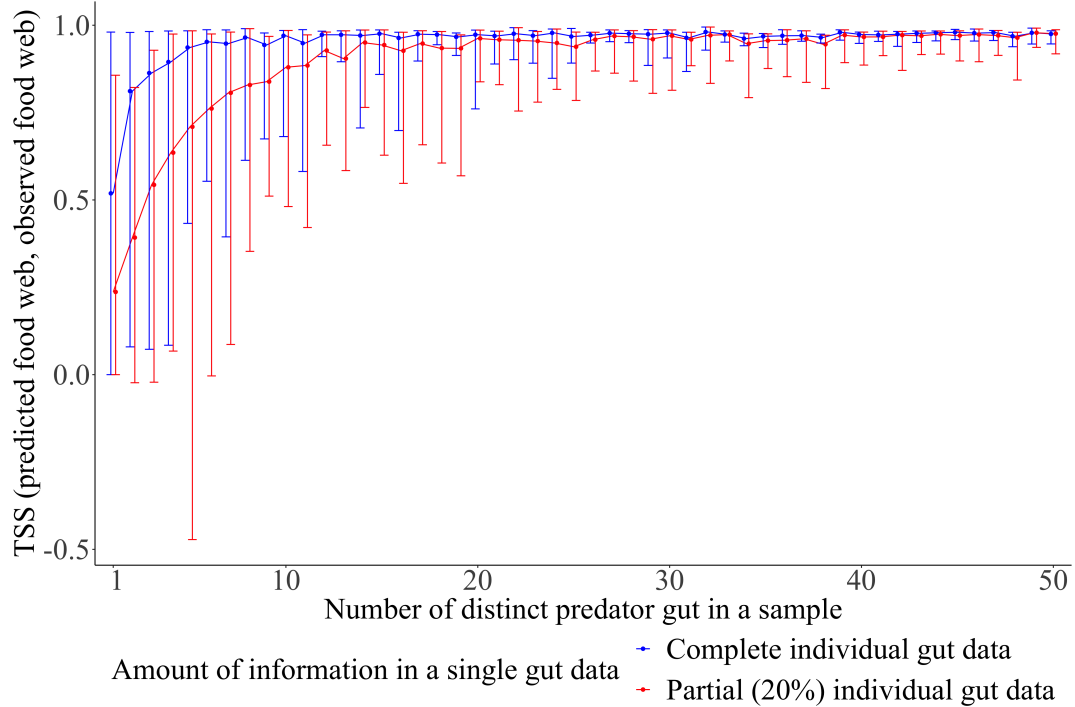## 7.1 Prediction using simulated gut content data from a simulated food web

Figure 7: True skill statistics between predicted food web and observed food web for a simulated small reef food web estimated for distinct predator guts in a sample. The observed simulated food web consists of 50 species and ... links. The vertical bars correspond to the prediction intervals of the true skill statistics with filled circles representing the corresponding mean. A prediction interval of the TSS is formed using a set of 100 accepted TSS values using the ABC method.

The true skill statistics (TSS) between the predicted food web and observed food web saturated with an increasing number of distinct predator guts in a sample (Fig. 7). The TSS of the predicted food webs estimated using the complete individual gut data had shorter prediction intervals resulting in less uncertainty, and higher mean TSS than that using the partial individual gut data. The maximum limit of the prediction interval of TSS estimated using the complete gut data and the partial gut data were almost equal, with the minimum limit of the prediction interval of TSS using partial gut data being lower than that from the complete gut data. Eventually, the gap between the mean TSS using the partial gut data and the complete gut data reduced with an increasing number of distinct predator guts suggesting when we have enough predator species' gut data, the achieved TSS was almost constant and hence independent of the amount of gut data.

The maximum TSS estimated using the complete gut data was very close to one and almost remained constant with an increasing number of different predator species sampled. With the gut data sample of only five distinct predator species, 95% of the maximum mean TSS was achieved when complete individual gut data was used, while the same was achieved with 15 predator species for partial gut data. This shows that one does not need to know the gut data of all the species to predict the food web and the accuracy is dependent on the completeness of an individual gut data.

8 References

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.

Berlow, E. L., J. A. Dunne, N. D. Martinez, P. B. Stark, R. J. Williams, and U. Brose. 2009. "Simple Prediction of Interaction Strengths in Complex Food Webs." *Proceedings of the National Academy of Sciences* 106 (1): 187–91. https://doi.org/10.1073/pnas.0806823106.

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.

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. https://doi.org/10.1046/j.1461-0248.2002.00354.x.

Hildrew, Alan G. 2009. "Chapter 4 Sustained Research on Stream Communities." In *Advances in Ecological Research*, 41:175–312. Elsevier. https://doi.org/10.1016/S0065-2504(09)00404-8.

Hyslop, E. J. 1980. "Stomach Contents Analysis—a Review of Methods and Their Application." *Journal of Fish Biology* 17 (4): 411–29. https://doi.org/10.1111/j.1095-8649.1980.tb02775.x.

Jordán, Ferenc. 2009. "Keystone Species and Food Webs." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1524): 1733–41. https://doi.org/10.1098/rstb.2008.0335.

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.

Layer, Katrin, Jens O. Riede, Alan G. Hildrew, and Guy Woodward. 2010. "Food Web Structure and Stability in 20 Streams Across a Wide pH Gradient." In *Advances in Ecological Research*, 42:265–99. Elsevier. https://doi.org/10.1016/B978-0-12-381363-3.00005-8.

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.

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

McCormack, Stacey A., Rowan Trebilco, Jessica Melbourne-Thomas, Julia L. Blanchard, Elizabeth A. Fulton, and Andrew Constable. 2019. "Using Stable Isotope Data to Advance Marine Food Web Modelling." *Reviews in Fish Biology and Fisheries* 29 (2): 277–96. https://doi.org/10.1007/s11160-019-09552-4.

McDonald-Madden, E., R. Sabbadin, E. T. Game, P. W. J. Baxter, I. Chadès, and H. P. Possingham. 2016. "Using Food-Web Theory to Conserve Ecosystems." *Nature Communications* 7 (1, 1). Nature Publishing Group: 10245. https://doi.org/10.1038/ncomms10245.

Peralta-Maraver, I., M. J. Lopez-Rodriguez, 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., A. P. Beckerman, J. O. Riede, and P. H. Warren. 2008. "Size, Foraging, and Food Web Structure." *Proceedings of the National Academy of Sciences* 105 (11): 4191–6. https://doi.org/10.1073/pnas.0710672105.

Rindorf, A, and P Lewy. 2004. "Bias in Estimating Food Consumption of Fish by Stomach-Content Analysis" 61: 12.

Roslin, Tomas, and Sanna Majaneva. 2016. "The Use of DNA Barcodes in Food Web Construction—Terrestrial and Aquatic Ecologists Unite!" Edited by Elizabeth Clare. *Genome* 59 (9): 603–28. https://doi.org/10.1139/gen-2015-0229.

Schmid-Araya, J. M., A. G. Hildrew, A. Robertson, P. E. Schmid, and J. Winterbottom. 2002. "The Importance of Meiofauna in Food Webs: Evidence from an Acid Stream." *Ecology* 83 (5): 1271–85. https://doi.org/10.1890/0012-9658(2002)083[1271:TIOMIF]2.0.CO;2.

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. 2010a. "Individual-Based Food Webs." In *Advances in Ecological Research*, 43:211–66. Elsevier. https://doi.org/10.1016/B978-0-12-385005-8.00006-X.

———. 2010b. "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.

Woodward, Guy, Dougie C. Speirs, and Alan G. Hildrew. 2005a. "Quantification and Resolution of a Complex, Size-Structured Food Web." In *Advances in Ecological Research*, 36:85–135. Elsevier. https://doi.org/10.1016/S0065-2504(05)36002-8.

———. 2005b. "Quantification and Resolution of a Complex, Size-Structured Food Web." In *Advances in Ecological Research*, 36:85–135. Elsevier. https://doi.org/10.1016/S0065-2504(05)36002-8.