Chapter 2: Effects of data composition and diversity on precision and accuracy of network inference

Contents

1 Abstract

2 Introduction

2.1 Knowing species interactions

Knowing who eats whom in a a food web can help us answer crucial ecological questions such as: what if the top predators are removed how would the food wed respond? which is/are keystone species? which species should be given more conservation effort? citreqd

2.2 Conventional methods of knowing interactions

Multiple approaches such as gut content analysis, stable isotope ratio compostion of tissues, experimentation have been used to know the predation links between species (citreqd). These approaches require a lot of sampling effort (citreqd).

2.3 Gut content analysis

Gut content analysis is performed to know what an individual has eaten and the items in the stomach are used to infer the diet of that species. There is considerable uncertainty involved in this method (Baker, Buckland, and Sheaves 2014). In fish, there is ambiguity while separating prey items in the gut. There are loose tissues which are not identifiable and cannot be assigned to a specific prey item with certainty. Even when it is possible to separate prey items and identify the prey, there are lot of 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 unquantifiable error which are difficult to interpret in the predator diet (Hyslop 1980; Rindorf and Lewy 2004).

2.4 Stable isotope ratios

Stable isotope ratios (SIR) of tissues from a single individual is used to quantify diet over varying temporal and spatial scales from one sample (Crawford, Mcdonald, and Bearhop 2008). Using SIR, 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.

2.5 Sampling effort problem

These methods of identifying require a lot of sampling effort (citreqd). These methods are quite tedious. It is diffcult to know how much to sample.

2.6 Research questions specific to gut content data

- Can ADBM predict the food web structure using partial gut content data?
  - first, we use a simulated food web and simulated gut content data to answer it.
  - second, we use real gut content data from Broadstone Stream data to answer it
- How does the result from above questions generalise over food webs?

2.7 Our approach

An important question in food web ecology is to understand how much data of prey-predator interactions is required to know the food web structure (citreqd). Is it possible to construct the food web by knowing only partial data? How should the sampling effort be distributed across species? Here, we present a theoretical study as well use a real data to answers these questions. We use a mechanistic model known as allometric diet breadth model (ADBM) which predicts food web structure. We provide partial data to the the model and observe if we can predict the food web structure and also quantify the accuracy of the prediction.

3 Materials and Methods

3.1 data synthesis and ADBM's predictions

First, we address these questions using *in silico* (Fig 1) by simulating food web structure from the allometric diet breadth model (ADBM) whose true parameters are known. We then synthesize data such as gut contents, stable isotope ratios, direct observation from the simulated food web. Next step would be to incorporate uncertainty in these data so that it mimics similar challenges as in real data. Then, we use these observed data which includes the uncertainty to predict the food web from the ADBM parameterised using approximate Bayesian computation (ABC). Here, we make use of the summary statistics in the rejection ABC approach.
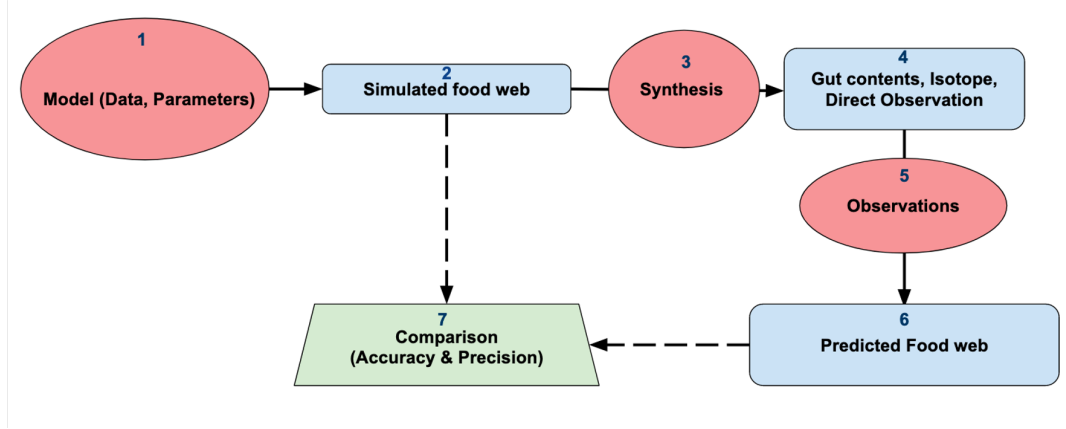

Figure 1:   Flowchart showing the method

3.2 Gut contents simulation

3.3 Stable Isotope Ratios simulation

3.4 Broadstone Stream example

- Broadstone Stream (51deg 05min N 0deg 03min E; 120 m above sea-level) is a second-order tributary of the River Medway in south-east England (see Hildrew (2009) for a detailed site description)
- The acidity of the stream (pH 4.7–6.6) excludes fish, resulting in an invertebrate-dominated food web (Woodward and Hildrew 2001). There are about 25 common invertebrate species (Woodward, Speirs, and Hildrew 2005). Among the common predators, there are three large species (Cordule- gaster boltonii Donovan, Sialis fuliginosa Pict. and Plectrocnemia conspersa [Curtis]) and three small species (the larvae of the tanypod midges Macrope- lopia nebulosa [Meigen], Trissopelopia longimana [Staeger], and Zavrelimyia barbatipes [Kieffer]), which we focus on in detail here when investigating size- structuring within the intraguild subweb.

4 Result

4.1 Prediction using simulated gut content data from a simulated food web

We investigated how the true skill statistics (TSS) between the predicted food web and observed food web varied with the number of predator species' sampled (Fig 2) for gut data. This investigation was done for simulated small reef food web using the ADBM. For every n number of predator species, we had 100 sets
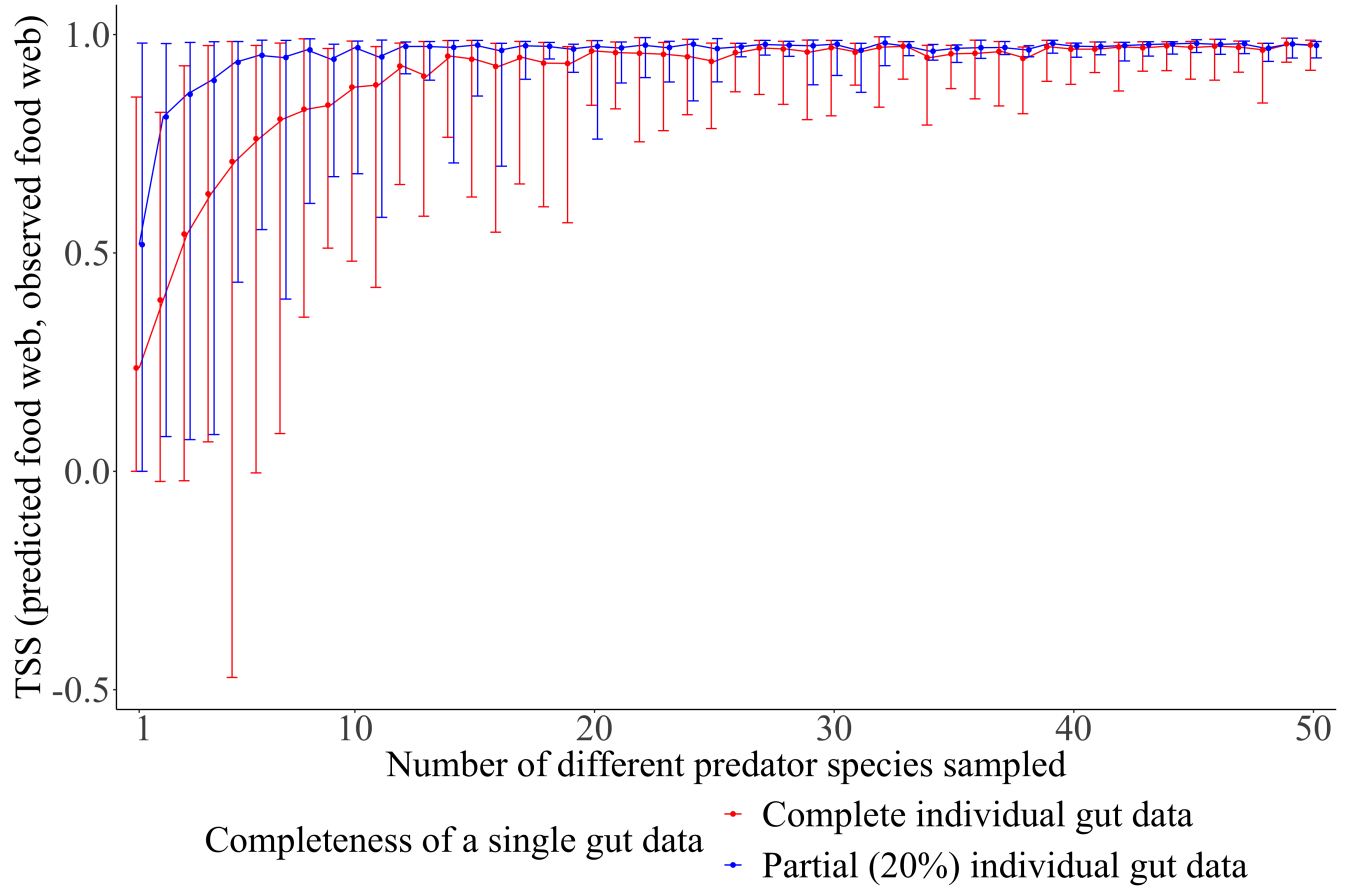
Figure 2: True skill statistics between predicted food web and observed food web for simulated small reef food web estimated for different number of predator species' guts sampled. For every value in the horizontal axis, there are 100 samples of randomly chosen species from a set of 50 species whose gut content data are simulated.

of gut data randomly selected from 50 species, where a set consists of n number of different predator guts. The TSS quickly saturated with increasing number of predator species. When compared with the partial individual gut data, the TSS estimated from that of the complete individual gut data had less uncertainty and higher mean TSS. The maximum values of TSS for complete gut data and partial gut data were almost equal, with the minimum TSS values for partial gut data lower than that of the complete gut data. Eventually, the gap between the mean TSS for partial gut data and complete gut data reduced with increasing number of predator species suggesting when we have enough predator species' gut data, the achieved TSS was almost constant whether the gut data of an individual predator was complete or not. The maximum TSS for complete gut data was very close to one and almost remained constant wih increasing number of different predator species sampled. With gut data of only five predator species, 95% of the maxiumum mean TSS was achieved for complete gut data, while with partial gut data the maximum mean TSS was achieved with 15 predator species. This shows that one does not need to know the gut data of all the species to predict the food web and the number of predator species is dependent on the completeness of an individual gut data. Hence, it is not necessary to know 100% accurate information of the gut data.

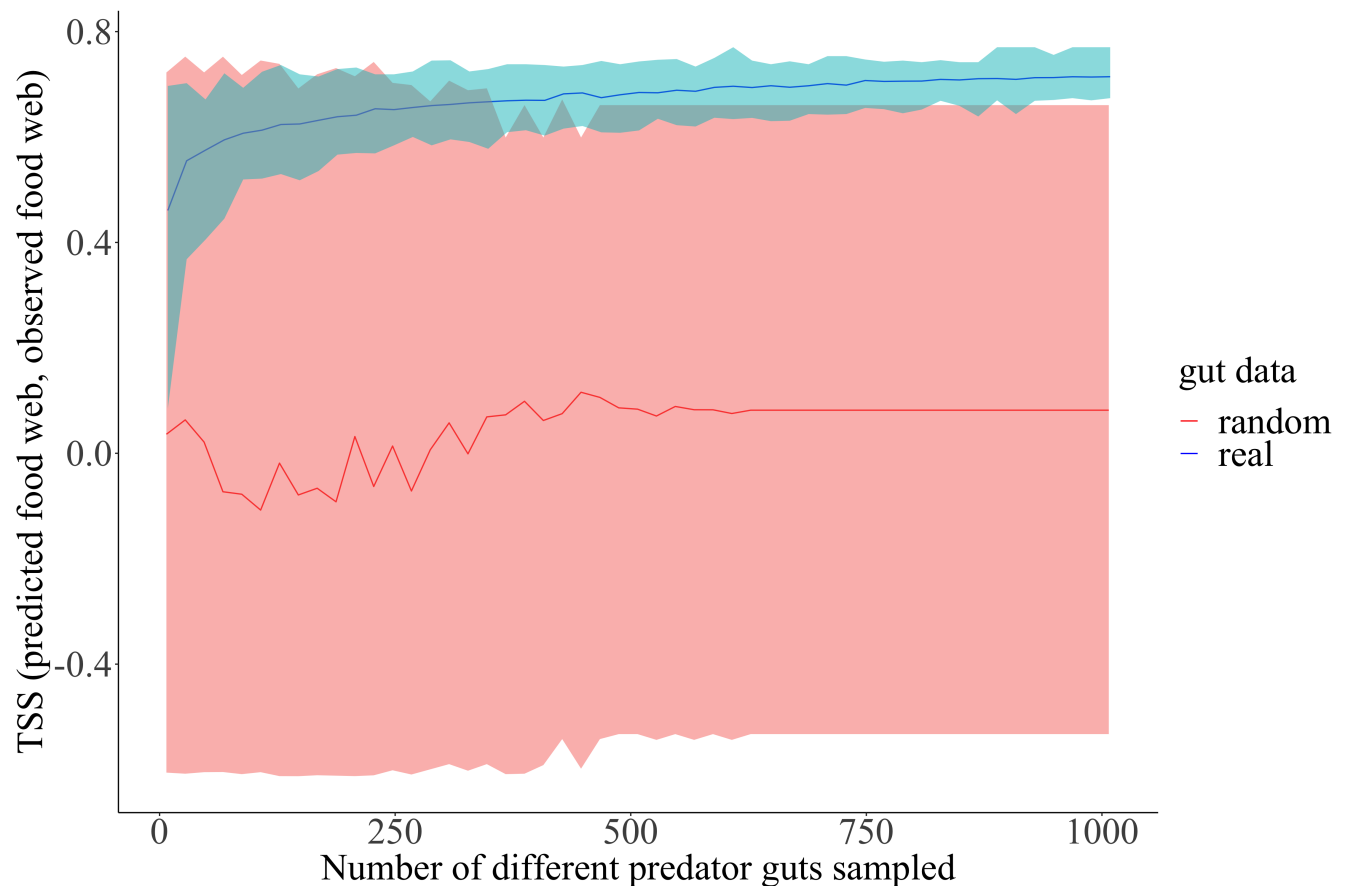## 4.2   Prediction using empirical (real) Broadstone stream gut content data

Figure 3:   True skill statistics between the predicted and observed Broadstone stream food web computed for number of different predator guts sampled (real gut data and randomly generated gut data). There were a total of 1008 predator guts. Blue curve corresponds to real gut data and red curve corresponds to randomly generated gut data. Line represents the mean and shaded region represents the range/uncertainty. Every point on the horizontal axis correspond to 100 samples.

The value of true skill statistics between the predicted food web and observed food web quickly saturated

with increasing number of sampled real predator guts (Fig 3). With only 228 number of real predator guts, 90% of the maximum mean TSS was achieved while 95% of the maximum TSS was achieved with around 428 real gut content data. The uncertainty (shaded region) in the TSS reduces with increasing number of predator guts. Although at low number of predator guts the maximum TSS is quite high but the uncertainty is high as well. This is quite possible just by chance which is suggested by the overlap of TSS region with that of randomly generated gut content data.

We compared the TSS predicted from real gut content data with randomly generated gut content data. Unsurprisingly, the range/uncertainty of TSS was high and almost remained constant with increasing number of predator guts sampled. The mean TSS fluctuated for lower number of predator guts sampled but eventually became constant, and was close to 0 which suggests the prediction is no better than random.

We investigated how the TSS varied with the number of predator guts sampled in fig 4 where the predator guts were divided into two cases depending on whether that predator has body size less than or greater than the average body size. We found a saturating curve in both the cases. The TSS corresponding to gut data less than average body size was higher than that corresponding to gut data greater than average body size for any number of predator guts sampled. Also, the TSS corresponding to the gut data less than average body size was higher when compared with the TSS corresponding to the gut data with all the body sizes.
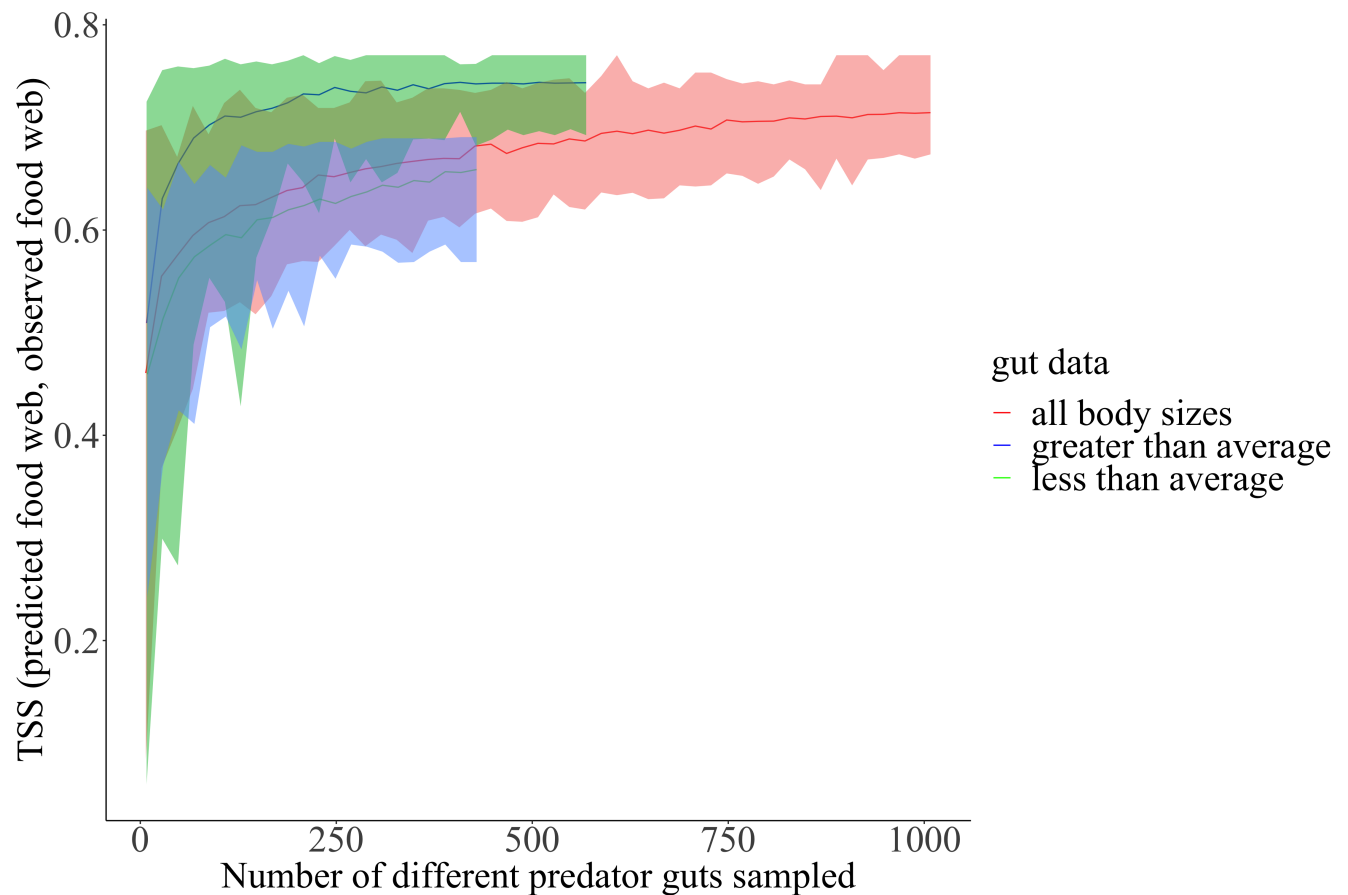

Figure 4: True skill statistics between the predicted food web and observed food web estimated with respect to the number of predator guts. The estimation is done for three sets of gut data: gut content data of predators less than the global mean body size, greater than the global mean body size and all the gut content data.

5 Discussion

5.1 Include digestion in gut content simulation

5.2 extrapolating ADBM's predictions to general food web

We first answer the questions with the help of ADBM. Now, the question is how much does it translates to real food web.

6 Conclusion

7 Appendix

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.

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.

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.

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

Woodward, Guy, and Alan G. Hildrew. 2001. "Invasion of a Stream Food Web by a New Top Predator." *Journal of Animal Ecology* 70 (2): 273–88. https://doi.org/10.1111/j.1365-2656.2001.00497.x.

Woodward, Guy, Dougie C. Speirs, and Alan G. Hildrew. 2005. "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.