Non-linear least squares model fitting of individual functional response curves

CMEE Miniproject report

Wenhua Zhou

CID: 00731413

wz2812@ic.ac.uk

Wordcount: 2616

Department of Life Sciences
Imperial College London
United Kingdom

March 6, 2020

1 Introduction

Functional response in ecology is the relationship between the intake rate of a consumer and the density of the target resource [1]. It describes in a way how a consumer responds to the changing density of its resource density. It is widely used in biological science to predict the change in predators number when the resource density varies by different conditions of the environment.

Functional response can also be applied to determine the stability of food webs in various ecosystems.

Functional response is first classified by C.S. Holling [2] into three basic types. After that, most of the researches in functional responses follow the work from Holling and the responses are classified as Holling's type I, II, and III functional response.

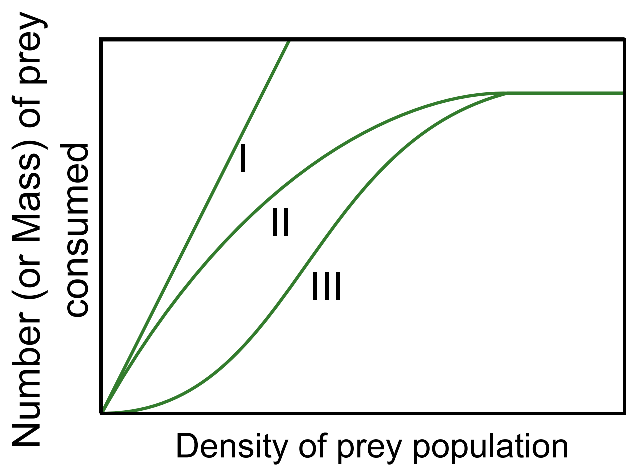

Figure 1: The general shape of Holling's type I, II, and III functional response curves

Type I functional response assumes proportional linear increase in consumption rate with resource density. And only when resource density increase to maximum, the consumption rate stops increasing and remains constant beyond that point. This implies a linear model

$$c = ax_R \tag{1}$$

where c is the consumption rate, a is the search rate, and x_R is the resource abundance.

A type II functional response is a non-linear model proposed by Holling on the basis of separating a consumer's time into two activities:'searching resource' and 'handling resource'[3]. As the consumption rate increase, more resource consumed requires more time to handle, which leads to a decrease in the total time for searching resource. Combining this with type I model leads to

$$c = \frac{ax_R}{1 + hax_R} \tag{2}$$

where h is handling time of the consumer for that resource. By this formula, saturation occurs at high levels of resource density, and consumption rate remains constant after that saturation point.

Type III functional response can be considered as a generalisation of the type II response to the form

$$c = \frac{a{x_R}^{q+1}}{1 + ha{x_R}^{q+1}} \tag{3}$$

with a extra parameter q to be a shape parameter for the response. The type III response is motivated by the learning behaviour of the consumers, while there is a learning time for them to improve their success rate in searching and attacking. This leads to a slow increase in the consumption rate at low resource density levels, and an acceleration as consumers meeting more resources until they fulfill their 'skills' and their success rate during searching reaches a maximum.

Apart from Holling's type responses, there are also some more mechanistic approaches for the functional response such as Beddington-DeAngelis functional response[4][5].

The main objective of the project is to accomplish non-linear least squares fitting for functional response curves using different models. By appropriate model comparison, describe the difference between the results and discussion the possible reasons. Then try to choose the best fitted model among them and describe how it can be improved reasonably.

2 Methods

2.1 Dataset

The dataset includes data collected through labs and field experiments across the world, discovering the behaviour between different consumer species and related resources species in decades. It has 4507 observations each with 68 different measurements including rates of consumption of a single resource species by a consumer species. Consumption rate and resource density in functional response are collected into this dataset as $N\_TraitValue$ and $ResDensity$ respectively. Each individual functional response curves can be identified by filtering the dataset with the same ID values. Alternative ways of identification are to reconstruct them by the same Citation or consumer/resource species ID.

$N\_TraitValue$ is the consumption rate in each observation, which is in terms of biomass quantity or number of individual resource consumed per unit time per unit consumer.

$ResDensity$ is the resource density, which is in terms of biomass quantity or number of individual resource per unit area or per unit volume.

ID is the ID number of individual consumer, there are 308 unique individual consumer in the dataset. The unique number of different citations and consumer/resource species ID are 113, 125, and 123 respectively. In addition, observations of each unique individual consumer are collected from the same citation with same consumer and resource species ID. Consumption rate unit and resource density unit also vary between different ID values. Therefore, fitting functional response curves individually tends to be more reliable than fitting the observations all together. Identifing by different citation and species ID are similar to ID number, but with less number of individuals. In this project, model fitting will be identified by ID number since fitting more individuals tends to be more precise and better in general.

2.2 Data preparation

In data preparation step, I write a python file *data_preparation.py* to read the dataset, and try to remove some poor observations which may affect model fitting results. In this case I just simply remove all observations with negative values and NAs in consumption rate and resource density, which is $N\_TraitValue$, $ResDensity$. Most of the measurements is not essential in my model fitting. In order to reduce data size, only three columns of the dataset with consumption rate, resource density, and individual ID number will be kept into the filtered dataset for model fitting.

2.3 Model introduction

Many mathematical approaches can be applied into functional responses problems, non-linear least squares method is one of the most well-known methods. Non-linear least squares regression is a form of non-linear regression using least squares analysis to fit m observations to a non-linear model with n unknown parameters ($m >= n$). It is widely used since it is fundamental and quick to fit [6].

Two main fields of interest in functional responses problems are consumption rate the resource density. From the simplest non-linear model to intermediate functional response model, a total of four different models will be fitted and computed. Further analysis is also required to test the performance of these models.

The first two models are the simplest non-linear approach: quadratic and cubic polynomial models

$$y = ax^2 + bx + c \tag{4}$$

$$y = ax^3 + bx^2 + cx + d \tag{5}$$

where y and x in functional responses problems are consumption rate and resource density. These models are trivial, but simple to fit and also essential in model fitting. Many of the complicated statistical models ends up with overfitting, simple models are widely used as a basic reference in model comparing. The rest models are the two classical functional response models mentioned previously. The Holling type II model

$$c = \frac{ax_R}{1 + hax_R} \tag{6}$$

and the generalised Holling model.

$$c = \frac{a{x_R}^{q+1}}{1 + ha{x_R}^{q+1}} \tag{7}$$

Generalised Holling model is a generalised form of the Holling type functional responses, with an additional parameter q which does not have any biological meaning. From the previous introduction of Holling's functional responses, when the handling time and shape parameter are both not considered ($h = 0$, $q = 0$):

$$c = ax_R \tag{8}$$

this is a Holling's type I functional response. When $h \neq 0$ and $q = 0$, generalised Holling model is equivalent to Holling's type II model. Also when q is positive, the consumption rate increases slowly at low resource density levels and then accelerates, which is exactly a Holling's type III response introducing a learning behaviour for the consumers.

Each model above will fit a functional response curve for each individual identified by different ID values of the data.

2.4 Calculate starting value

The main challenge in NLLS fitting is finding the right starting values. A unique starting value need to be determined for each individual functional response for each model. Algorithms are essential to generate large number of functional responses for parameters in each model.

In quadratic and cubic models, the starting value can be simple since they are likely to fit in most cases. The parameters in these models are the coefficients of the polynomials a, b, c, and d. In model fitting step, I consider all the starting value of these parameters to be a random normal distribution, with mean to be the same as the mean in each individual and variance 1.

Holling's type II model has two parameters, the consumer's search rate a and the resource handling time h. Similar approach has been made to search rate by Samraat et al.[7], using machanistic model to define search rate of the functional response in terms of body size of consumers and resources

$$a = s_D v_r d^{D-1} \tag{9}$$

with v_r and d to be the searching speed and searching radius. However, in small sample size in each individual this model is highly likely to overfit. In addition, there is no precise data and information about body size and searching radius in our dataset, so I find out the starting search rate a from the data by calculating the steepest gradient of the consumption rate increase with resource density. The consumption rate saturates at high levels of resource density. From the equation (2) of the Holling type II model, when x_R goes to ∞ the consumption rate will tends to be $1/h$. Therefore, the handling time h can be determined by the asymptotic consumption rate. More precisely, the asymptotic consumption rate is simply considered as

$$c_{asy} = c_{max} + \frac{1}{2}(c_{max} - c_{\text{second largest}}) \tag{10}$$

that implies

$$\begin{aligned} h &= \frac{1}{c_{asy}} \\ &= \frac{1}{c_{max} + \frac{1}{2}(c_{max} - c_{\text{second largest}})} \end{aligned}$$

Similar to Holling type II model, the generalised Holling model have the same parameters a and h. The additional parameter q determines the dimension of the resource density. The starting values of q is set to be a random uniform distribution from 0 to 3. This value satisfies a clear separation from the Holling's type II model and also prevents fitting failure when the dimension of the resource density is too large.

2.5 Model fitting

Once the starting values and data are all set, model fitting process will not be difficult. The main fitting step is accomplished in R using function $nlsLM$ in package minpack.lm[8]. This function is a modified version of non-linear fitting which is available for most of the generic function used in model analysis(e.g AIC, BIC). The model with the most parameter is the cubic polynomial model with 4 different parameters. By definition of non-linear least squares requires at least 4 observation to fit. In my model fitting step, I only consider the ID values with 5 or more observations which will not cause bad fitted results due to lack of observations.

For a total of 308 different individuals' data, the four models are fitted every time with unique calculated starting values of their parameters. Some mechanistic models are unlikely to fit for all individuals' data, so it is essential to use $tryCatch$ to prevent error popping up during model fitting. However, about 40% of the individuals is going to fail in at least one model fitting by this method. To increase the fitting success rate, different starting values are worth trying. Given some of the starting values of parameters in the model are determined by a distribution which varies every time, we can keep the fitting process until it is fitted correctly. Using this technique, a upper bound number 20 is set for each fitting loop. Also for parameters a and h with fixed starting values in each individual, random numbers close to the fixed values can be considered within each loop. I accept a random number within 20% error of the fixed starting value for both a and h in the model fitting. Using the parameters of fitted models, we can predict the fitted observations of the dataset and record the predicted consumption rate in each observation for each model for future plotting. Statistical measures AIC, BIC and R^2 are calculated for model comparison and analysis.

2.6 Computing tools

There are two shell scripts used in the project: one to compile the pdf from the tex file, and a bash script runs the whole project and all other python and R files.

Python is used in data preparation step and the only package used is *pandas* for reading and writing csv dataset.

R is the main computing tool in this project. NLLS fitting, plotting and analysis are all done in R within different files. In NLLS fitting, *minpack.lm* is the main package used for model fitting. *ggplot*2 is the main package used in plotting, it is a great tool to visualise the data in R.

3 Plotting and results

Using non-linear least squares fitting, 4507 observations of consumption rate are all fitted with four different models. Each individual with correctly fitted models can plot a prediction line for the fitted curve using *ggplot*.

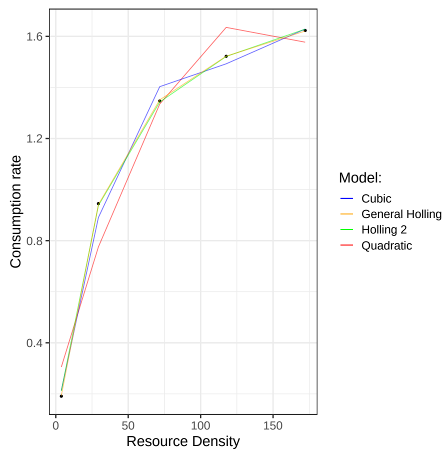

Figure 1: Need a finished plot with ID number

However, not all individuals can be fitted nicely by all the models. Lack of observations for each individual, outliers and bad observation results are the main reason causing failure in fitting. In the end among 308 individuals, 13 individuals are filtered out due to lack of observations.

	Fitted models			
Individuals	Quadratic	Cubic	Holling's Type II	Generalised Holling
Fitted	295	294	234	231
Total	295	295	295	295
Success Rate	100%	99.6%	79.3%	78.3%

Table 1: Number of individuals fitted and the success rate for each model in model fitting

The fitting table for the rest of 295 individuals implies that quadratic and cubic polynomials are simple to fit with a extremely high percentage of success rate. On the other hand, models of functional responses share similar success rate, while generalised form of Holling's functional response are only slightly harder to fit than the more specific type II model.

4 Analysis and discussion

To examine the results in a statistical way, AIC, BIC and the coefficient of determination (R squared) for all individuals are recorded during the model fitting process. The Akaike information criterion (AIC) is an estimator of data, describing the quality of the in statistical modelling[9]. The value of AIC is determined by number of parameters and the maximum value of the likelihood function with the formula

$$\text{AIC} = 2k - 2ln(\widehat{L}) \tag{11}$$

Similarly, the Bayesian information criterion (BIC)[10] is defined as

$$\text{BIC} = ln(n)k - 2ln(\widehat{L}) \tag{12}$$

where k is the number of parameters, $\widehat{L}$ is the maximum likelihood, and n is the number of observations in data. Both of them deal with the trade off between the goodness of fit and the simplicity of the model. The sample size for each individual is relatively small in my data. AIC value will not change with the number of observation in data, but it has a modification for small sample sizes AICc with

$$\text{AICc} = \text{AIC} + \frac{2k^2 + 2k}{n - k - 1} \tag{13}$$

In this project I will use AIC to analyse the performance of models and compare that with AICc and discuss the results. The model with smaller AIC value is the better fit, so for each individual I find the best model with the minimum AIC value and models with AIC values difference less than 2 from the minimum AIC will all considered as they are equally the best. At last, count the total number of best fit for each model and compare results.

	Fitted models			
AIC	Quadratic	Cubic	Holling's Type II	Generalised Holling
Best fit	146	220	93	138
Total fit	295	294	234	231
Number of parameters	3	4	2	3
Best fit rate	49.5%	74.8%	39.7%	59.7%

Table 2: Best fit rate among models using AIC

The table above illustrates two main problems in our case when applying AIC: the best fit rate are highly related to the number of parameters, which AIC is highly likely to overfit. Also some individuals can also fitted by quadratic and cubic model, which makes best fit in these individuals less competitive. Therefore, I improve the table using AICc instead, and only find the best fit among those individuals can be fitted by all of the models.

	Fitted models			
AICc	Quadratic	Cubic	Holling's Type II	Generalised Holling
Best fit	69	43	145	78
Total fit	231	231	231	231
Number of parameters	3	4	2	3
Best fit rate	29.8%	18.6%	62.8%	33.8%

Table 3: Best fit rate among models using AICc

The best fit rate changes completely after using AICc instead of AIC. It turns out cubic model is just overfitting with small sample sizes and the Holling's type II is the best model among them with almost double the best fit rate of the generalised Holling model.

R square value is the proportion of the variance in the dependent variable from independent variable. It varies from $-\infty$ to 1, and the closer it gets to 1, the better the fitting. However, R square values are not often used in determine the goodness of fitted models since they do not indicate whether a regression model is adequate. For 4 different models, the averaging R square values of cubic polynomial is the highest, following by quadratic, generalised Holling's response and Holling's type II response.

This is also because the various number of parameters. Therefore R square is not a good reference for testing the goodness of fit in this case.

From my analysis comparing these four models, the Holling's type II model tends to be the best model among them. Cubic polynomial fits nice curves for most of the individuals, but overfit badly by my AICc results. Generalised Holling model should be better in a large sample size compared to Holling type II model, but the shape parameter is not really necessary in this dataset. Due to lack of sample size in most of individuals, it is difficult to apply more parameters and more complicated models. In my case, average sample size for each individual which can be fitted for all models is only less than 9. If there are more samples for each individual, there is a huge chance that Holling's type II model fitting will be overtaken by generalised Holling model or the Holling's model with S. Pawar's model[7] for the search rate.

Bibliography

[1] M. Solomon, "The natural control of animal populations," *The Journal of Animal Ecology*, pp. 1–35, 1949.

[2] C. S. Holling, "The components of predation as revealed by a study of small-mammal predation of the european pine sawfly," *The Canadian Entomologist*, vol. 91, no. 5, pp. 293–320, 1959.

[3] J. Dawes and M. Souza, "A derivation of holling's type i, ii and iii functional responses in predator–prey systems," *Journal of theoretical biology*, vol. 327, pp. 11–22, 2013.

[4] R. S. Cantrell and C. Cosner, "On the dynamics of predator–prey models with the beddington–deangelis functional response," *Journal of Mathematical Analysis and Applications*, vol. 257, no. 1, pp. 206–222, 2001.

[5] S. Geritz and M. Gyllenberg, "A mechanistic derivation of the deangelis–beddington functional response," *Journal of theoretical biology*, vol. 314, pp. 106–108, 2012.

[6] F. Cadet, S. Garrigues, and M. de la Guardia, "Quantitative analysis, infrared," *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, 2006.

[7] S. Pawar, A. I. Dell, and V. M. Savage, "Dimensionality of consumer search space drives trophic interaction strengths," *Nature*, vol. 486, no. 7404, pp. 485–489, 2012.

[8] T. V. Elzhov, K. M. Mullen, A.-N. Spiess, B. Bolker, M. K. M. Mullen, and M. Suggests, "Package 'minpack. lm'," *Title R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds'. Available at: https://cran. rproject. org/web/packages/minpack. lm/minpack. lm. pdf*, 2016.

[9] Y. Sakamoto, M. Ishiguro, and G. Kitagawa, "Akaike information criterion statistics," *Dordrecht, The Netherlands: D. Reidel*, vol. 81, 1986.

[10] S. Chen, P. Gopalakrishnan *et al.*, "Speaker, environment and channel change detection and clustering via the bayesian information criterion," in *Proc. DARPA broadcast news transcription and understanding workshop*, vol. 8. Virginia, USA, 1998, pp. 127–132.