

Miniproject Report

Amelia Wake

03.12.21

Evaluation of applied mathematical models in population biology, using food microbiology as a case study.

Abstract

This paper investigates the ability of different mathematical models to fit functional responses data across species, more specifically it contrasts and compares mechanistic and phenomenological models against each other.

1 Introduction

The fundamental nature of biology is complex. As increasingly vast amounts of sophisticated data regarding the biological world becomes more readily available through the evolution of technology, more complex mechanisms and tools are required to explain biological phenomena (Bolker et al., 2013). Null hypothesis-testing becomes ever more obsolete as researchers adopt more intricate and robust approaches to exploring data (Johnson et al., 2004). Model selection, based upon likelihood theory, is quickly becoming the most popular method in certain fields, such as microbiology (Ferrer et al., 2009). This approach allows for the handling of irregular, nonlinear, heteroscedastic data that was unsuitable for the simple ANOVA analyses of the past (Bolker et al., 2013). Biology is becoming an increasingly inter-disciplinary field as the incorporation of mathematics, computer science and physics becomes more accepted (Gunawardena, 2014). Many biological phenomena can be fitted to differential equations, and in this way applied mathematics gives us the tools, via mathematical models, to understand complexity (Transtrum et al., 2016). Mathematical models can not solely express the complexity of a biological system, however they provide a good representation when they are intricate enough to represent the system, while being

transparent enough to allow for practical inferences and predictions across variable conditions (Transtrum et al., 2016). Model selection has three main advantages over null hypothesis-testing. It allows for multiple rival models to be evaluated against a data set simultaneously, instead of against an arbitrary probability threshold. The degree to which a model fits the data, the more support it lends to the associated hypothesis. Additionally, weight-based rankings provide a quantitative method of measuring support to said hypothesis. Finally, in cases where models have indistinguishable or nearly equal AIC (Akaike Information Criterion) values, model averaging allows us to combat this problem by ensuring robust parameter estimates and predictions (Johnson et al., 2004). Mathematical models have had a significant impact on furthering our understanding of experimental observations, as well as biological systems and mechanisms, for the last twenty years (Jin, 2017).

2 Materials & Methods

1. Data wrangling

A data set containing measurements of change in biomass and number of cells of microbes over time was prepared for model fitting by creating unique IDs, in order to allow for the extrapolation of unique datasets. These unique IDs were created by concatenating unique temperature-species-medium-citation-replicate combinations to produce single population growth rate curves. Missing and problematic variables were stripped from the data in order to not skew the model fitting and analyses.

2. Model fitting – cubic polynomial, logistic, and Gompertz models.

In order to provide a solution for the question posed in this research project I.e. How well do different mathematical models fit to functional responses data across species? A cubic polynomial regression was chosen as the phenomenological model to be evaluated against two mechanistic models: logistic and Gompertz models. Two mechanistic models were chosen instead of one due to the limiting factors of the logistic model approach in regards to fitting to population growth data. The logistic model does not take into consideration the ‘lag time’ of microbial populations before the ‘exponential growth’ phase, however the Gompertz model considers this and produces a more appropriate sigmoidal curve.

3. Final plotting and analysis script

AIC (Akaike Information Criterion) values were used as the main unit of model comparison, due to r-squared values not providing us with much

useful data for the mechanistic models. However, r-squared values were still considered when qualitatively comparing the fit of each model. All three models – cubic polynomial, logistic regression and Gompertz models were plotted on one graph per unique ID, in order to allow for efficient comparison of goodness of fit to the data.

A foundational equation of population biology is:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \quad (1)$$

It was first proposed by Verhulst in 1838 [?].



References