



# The Relative Strengths of Linear and Non-Linear Models for **Viewing** Microbial Growth across Individual Populations

Dec 2021

Iw121



# 1 Introduction

Use of mathematical models has expanded ecology's research frontiers beyond what is immediately observed, raising new questions about modality, reality and our immediate explicanda. One particularly salient distinction, is that of Phenomenological versus Mechanistic models. In line with the popperian tradition of knowledge first by hypothesis and then evidence, Mechanistic models aim to explain a given phenomenon in virtue of its discrete causal mechanisms. These models specify sets of parameters and relationships between them thought to reflect the interacting parts which give shape to a given biological phenomenon (hypothesis). The strength of these parameters for explaining a given phenomenon, and the relative contribution of each value can be then ascertained in virtue of how well the parameters fit a given data set (evidence).

As highlighted by Levins, however, natural (and particularly biological) phenomenon are influenced by such a wide range of factors and interactions that finding a single Mechanistic model which is both meaningful and sufficiently representative of a data set is difficult to come by/citelevins1966strategy. Levins articulates this as an inevitable trade off between "Precision", "Realism", and "Generality". Mechanistic models which aim to be highly realistic and precise do so at the cost of generality- explaining phenomenon to a high degree of accuracy, but in very specific settings.

Phenomenological models on the other hand are aimed at discovering patterns of interaction over and above any consideration of causal mechanisms. They are concerned with finding patterns of interactions within a pre-specified, homogeneous data set, rather than objects in the world which are messy and easily fragmented. As opposed to Mechanistic models, these tend to have a high degree of generality, making them useful predictors, and perhaps explaining the increasing trend towards phenomenological models in biology/citeBUCHANAN1997313.

In this paper, I will consider the use of two linear (Phenomenological) and one non-linear (Mechanistic) model for understanding patterns of microbial growth across different populations.

## 2 Method

### 2.1 Data

The datasets used are a collection of Microbial Growth records from various labs across the world. Our main focus for model fit is the pattern of change in microbial biomass/population size (PopBio) over time (Time). A simple linear model could be fitted from Time-start to Time-finish to give the average Growth rate of a population. However, bacteria do not tend to regenerate at a stable rate, rather Growth can be more accurately characterised by three distinctive "Lag" "Exponential" and "Stationary" phases, creating a pattern more akin to a sigmoid curve.



## 2.2 Models Model fitting

To best approximate this curve, I fit two simple polynomial linear models (Quadratic and Cubic) using the formulas below, as well as the Gompertz model which explicitly aims to predict the phases of cell growth, taking the end of the lag phase (t<sub>lag</sub>), exponential growth phase (maximum growth rate, r<sub>max</sub>) and approaching values (max abundance) as its parameters.

The two linear models can be expressed as such, where x is Time and y is our Population size at T.

Gompertz model is demonstrated below, where "t<sub>lag</sub>" is the end of the growth phase, "r<sub>max</sub>" is maximum growth rate, K is carrying capacity (maximum population size) and "N<sub>0</sub>" is the initial population size.

In line with expectations of the Gompertz model, I have log transformed PopBio in the data sets for both models. This will also better amplify model deviations from the datasets.

In order to assess which model best describes microbial growth across the records, I subset the data into individual microbial populations (growth curves) and fit each model to each distinctive population.

For the Gompertz model, I used nlsLM() in R (a more robust nonlinear least squares package which uses the Levenberg-Marquardt algorithm instead of the traditional Gauss-Newton algorithm) to optimise model fit to each data set. This involved first generating viable starting values for each parameter, then fitting them to the Gompertz model and using the coefficients from the model of best fit as universal starting values.

Once fitted, I assessed best fit via model selection criterion, using the Akaike Information Criterion (AIC) to generate maximum likelihood scores for each model in each sample population. I choose to use AIC as it aims to maximise fit while preserving simplicity (thus it penalises for too many additional parameters). Thus it encounters the tensions earlier of generality versus approximation of reality in decided how well a model describes a given phenomenon of interest[?].

The best model per data set was then selected according to the model with the smallest AIC. Lastly, the models were plotted over each data set to visually assess fit and make adjustments where needed.

## 2.3 Computing

I chose to conduct the modelling and data analysis for this project in R, to make use of its comprehensive tools for data manipulation, visualisation and statistical testing.

| Model     | Count<br>"Best fit" |
|-----------|---------------------|
| Cubic     | 105                 |
| Gompertz  | 143                 |
| Quadratic | 29                  |

Table 1: Table showing number of times ("count") each model has highest AIC for a given data set (unique population)

### 3 Results

#### 3.1 Model selection: Best fit

The table below shows the sum of 'best fits' according to AIC for each model, as well as the minimum, maximum and mean AIC values for each model across the different data sets.

#### 3.2 Visualisations



The Charts below show three datasets with the models overlaid.

### 4 Discussion

As seen in the above table, the Gompertz nonlinear model fit best to the data sets at hand, converging to the strongest fit out of the three models (AIC) 143 times out of 277 datasets.