miniproject write up

kate.griffin21

December 2021

Contents

1 Introduction

Food spoilage may be defined as any change in a food product which deems it unsuitable for consumption by a consumer [1]. The most common cause of food spoilage is that caused by microbes, which can alter a food item's flavour, smell, texture and taste [1]. The surfaces of food products are often nutrient rich and, in combination with suitable environmental conditions, this can facilitate the growth of biofilms [1]. In the food industry, predicting microbial growth may be useful in evaluating risk and shelf life of perishable goods [3][1][8]. Microbial growth rates may be measured by plotting population density against time on a graph and fitting a straight line through the exponential growth phase. The slope of this line represents the maximum growth rate (r max). In recent years, predictive microbiologists use mathematical models which can describe the growth of microorganisms along the sigmoidal bacterial growth curve. Such models include the Gompertz (cite), Logistic (cite), Baranyi and Buchanan models[3].

The growth of a bacteria in batch culture may be divided into five main phases based on the slope of its growth curve [4][3][2][8]. The first phase is the lag phase, where bacteria adapt themselves in preparation for their growth. During this phase, a number of genes are activated, including those involved in the uptake of nutrients, DNA repair, RNA processing, cell division and other biomolecular processes. The next phase is the exponential growth phase, where cell division occurs at a constant rate following a geometric progression. Once carrying capacity of the environment (i.e., growing media) has been reached, the stationary phase commences. During this phase, bacteria stop replicating. The next phase is the death phase, which occurs when the bacterial cells lose viability. The last stage is the long-term stationary phase, where bacteria can remain for long periods of time without nutrients [2].

Ecological and evolutionary models may be considered as a spectrum between phenomenological to mechanistic methods [7]. Models may be described as mechanistic if they include a number of parameters with biological definitions, which may be measured independently of the observed data [6][7]. In comparison, phenomenological models (e.g., cubic and quadratic polynomial models) measure a hypothesised relationship between two or more variables in a dataset, which can be fitted to the data [7]. When applied to measure population growth, the logistic model takes the carrying capacity of the environment into account. It demonstrates that the per capita growth rate decreases as population density increases [5]. This is because populations grow exponentially while their population size is low and resources are not limiting, and then slows down as the population becomes so large that resources become limited (i.e., the Malthusian principle). The logistic model may be considered a mechanistic model when modelling population size, as it can be modified to include parameters which have biological meaning.

Although mechanistic models are more robust than phenomenological models, they are not necessarily more valuable in making predictions [7]. Mechanistic models are particularly suitable for analysing populations when predictions outside of the known set of parameters is required and there is also enough existing data to make a priori estimates of its parameters [7]. When a biological system or processes can be explained by relatively few variables, a simpler model with fewer parameters may be more useful [6].The aim of my analysis is to address the question as to whether phenomenological models or logistic models are better for modelling population growth in microbial batch cultures. In my analysis I will be comparing a quadratic polynomial model to a mechanistic logistic model in predicting the change in microbe abundance over time.

2 Methods

2.1 Preparing the data

The dataset used in this analysis contains measurements of biomass/number of cells of forty-five species of microbes grown on eighteen different media over time. The data was collected from ten different studies across the world. Field names are defined in the metadata file. The two main variables I will be comparing are "PopBio" (abundance) and "Time". Single population growth curves were identified as unique combinations of temperature, growing media, species and citation. These unique IDs were saved in a new column and used to subset the data. The data was wrangled in Python. Firstly, the data was checked for NA values as well as negative values for PopBio and Time. Rows which contained negative values were deleted. There were no NA values found in the dataset. The wrangled data with the unique ID column was then saved as a .csv file.

2.2 Fitting the models

The models were defined and ran in R. The models were fit in order to find their estimated parameters. AIC and BIC were then calculated to find the fit of the model parameters to the observed data. The first model fitted was the polynomial quadratic model, the equation for which is:

$$B = B_0 + B_1 x + B_2 x^2$$

where x represents an independent variable, which in this analysis is time. B_0, B_1, B_2 represents the parameters, i.e., the coefficients of the equation (respectively: the quadratic coefficient, the linear coefficient and the constant or term). Once the model was defined, it was ran on each subset of data. The AIC and BIC were calculated and saved.

The next model to be fit was the logistic model, the classic equation for which is:

$$Nt = \frac{N_0 K e^{rt}}{K + N_0(e^{rt} - 1)}$$

The three parameters defined in the model are: r (maximum growth rate, also known as R max), K (carrying capacity) and N_0 (initial population). Nt represents the population size at time t, and No is the initial population size. As with the quadratic model, the model was fit to find its parameters and ran on each subset. AIC and BIC were calculated to evaluate the fit.

2.3 Analysing AIC and BIC

In order to compare the models, the difference in AIC between the quadratic and the logistic model was calculated for each subset. If the absolute

difference was greater than 2, then the model with the smaller AIC value was selected as the better model. I created a tally of the models selected for each subset to compare which model was selected more often based on AIC. BIC was also then compared using the same method as AIC.

References

[1] Young-Min Bae, Ling Zheng, Jeong-Eun Hyun, Kyu-Seok Jung, Sunggi Heu, and Sun-Young Lee. Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce. *Journal of food science*, 79(10):M2072–M2080, 2014.

[2] Steven E Finkel. Long-term survival during stationary phase: evolution and the gasp phenotype. *Nature Reviews Microbiology*, 4(2):113–120, 2006.

[3] Lihan Huang. Optimization of a new mathematical model for bacterial growth. *Food Control*, 32(1):283–288, 2013.

[4] Matthew D Rolfe, Christopher J Rice, Sacha Lucchini, Carmen Pin, Arthur Thompson, Andrew DS Cameron, Mark Alston, Michael F Stringer, Roy P Betts, József Baranyi, et al. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of bacteriology*, 194(3):686–701, 2012.

[5] A Stefan, H Geritz, and Éva Kisdi. Mathematical ecology: why mechanistic models? *Journal of mathematical biology*, 65(6-7):1411, 2012.

[6] Mark K Transtrum and Peng Qiu. Bridging mechanistic and phenomenological models of complex biological systems. *PLoS computational biology*, 12(5):e1004915, 2016.

[7] Craig R White and Dustin J Marshall. Should we care if models are phenomenological or mechanistic? *Trends in ecology & evolution*, 34(4):276–278, 2019.

[8] MH Zwietering, Il Jongenburger, FM Rombouts, and KJAEM Van't Riet. Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–1881, 1990.