

Imperial College London

Comparison of Models for Fitting Population Growth Curves

Yuqing Zhou

Department of Life Sciences, Faculty of Natural Sciences,
Imperial College London

March, 2020

Approximate Word Count: 1500

Comparison of Models for fitting Population Growth Curves

Yuqing Zhou (CID: 01807289)

Abstract

Several non-linear models(logistic, Gompertz, Baranyi, Buchunana) were compared by assessing the fits of the models for an emperical dataset.

Keywords: Model selection, microbial population growth rate, model Comparison, logistic growth, Gompertz model, Baranyi model

1 Introduction

Linear and nonlinear regression analysis has become an essential tool to analyze biological data and make biological inferences. Models have been developed to allow the description of observed biological patterns in general ways which provide insight into responsible factors (Johnson and Omland, 2004). The approach of model selection offers a quantitative way to measure relative support for a set of competing hypotheses represented by models, which has become a preferred alternative to null hypothesis testing approach to answer ecological and evolutionary questions (Hilborn and Mangel, 1997).

One of the widely accepted modeling dichotomies illustrated is mechanistic (process) and phenomenological (pattern) models (Bolker, 2008). The phenomenological model aims to quantify experimentally observed patterns and is not derived from mechanistic considerations, while the mechanistic model concerned more with the underlying processes based on theoretical expectations. The use of mathematical models is increasingly employed in microbiology to describe and predict the behaviour of microorganisms, which bears tremendous hopes in application in field of food microbiology and other areas (e.g. Baranyi and Roberts, 1995).

Typical microbial growth in a closed habitat, e.g. batch culture follows a four distinct stages including lag phase, exponential growth phase, stationary phases and mortality phase (McKellar and Lu, 2004). During the lag phase the cells are assumed to start a growth rate of zero from initial population size (N_0) and prepare for growth before beginning exponential growth,

resulting in a lag time (t_{lag}). During the exponential phase, cells accelerate to a maximum growth level (N_{max}). The maximum growth rate(r_{max}) is traditionally defined by the slope of the straight line fitted in the exponential phase. During the stationary phase, the growth rate slows as the population size nears carrying capacity, and then the number of cells in the culture stabilises.

The majority of the developed models found in literature do not consider the mortality phase, such as the Gompertz models (Gibson et al., 1988), the Baranyi model (Baranyi and Roberts, 1995), the Buchanan model (or the three-phase logistic model, Buchanan et al., 1997), and logistical model (Ricker, 1979).

Various models devised hitherto contain the four parameters mentioned above (N_0 , t_{lag} , N_{max} and r_{max}). In this work, the phenomenological models including cubic polynomial models, the logistic model with t_{lag} , the modified Gompertz models (Zwietering et al., 1990), the Baranyi and the Buchanan model, and mechanistic model - the logistic model (does not contain t_{lag}), were analysed and evaluated. The objective of this work is to address the question of how well do different mathematical models fit to growth curve across species.

2 Methods

2.1 Fitting Model

Five model, the cubic polynomial models, the logistic model, the shifted logistic model, the Gompertz model, the Baranyi model and the Buchanan model were compared.

Cubic polynomial model (Eqn 1.)

$$N = N_0 + N_1T + N_2T^2 + N_3T^3 \quad (1)$$

The cubic polynomial model describes changes in population size (N) at any given time (T). Time is measured in hours. None of the parameters have any biological meaning in this case.

Logistic Growth Model(Eqn 2.)

$$N_t = \frac{N_0N_{max}e^{rt}}{N_{max} + N_0(e^{rt} - 1)} \quad (2)$$

N_t is population size at time t , N_0 is initial population size, r is maximum growth rate (AKA r_{max}), N_{max} is maximum population density (AKA carrying capacity).

Logistic Growth Model with a Lag Time (Eqn 3.)

$$N_t = \frac{N_0 N_{max} e^{rt}}{N_{max} + N_0(e^{rt} - 1)} \quad (3)$$

N_t is population size at time t , N_0 is initial population size, r is maximum growth rate (AKA r_{max}), N_{max} is maximum population density (AKA carrying capacity).

Gompertz Model(Eqn 4.)

$$N_t = A e^{-e^{\frac{r_{max} e(t_{lag} - t)}{A} + 1}} \quad (4)$$

$$A = \ln\left(\frac{N_{max}}{N_0}\right) \quad (5)$$

The maximum growth rate (r_{max}) is the tangent to the inflection point, t_{lag} is the x-axis intercept to this tangent (duration of the delay before the population starts growing exponentially) A is the asymptote, N_0 is initial population size, N_{max} is maximum population size.

Baranyi Model(Eqn 5.)

$$N_t = N_0 + r_{max} A_t - \ln\left(1 + \frac{e^{r_{max} A_t} - 1}{e^{N_{max} - N_0}}\right) \quad (6)$$

$$A_t = t + \frac{1}{r_{max}} \cdot \ln\left(\frac{e^{-r_{max} t} + h_0}{1 + h_0}\right) \quad (7)$$

$$t_{lag} = \frac{\ln(1 + \frac{1}{h_0})}{r_{max}} \quad (8)$$

The Baranyi model introduces a new dimensionless parameter h_0 which represents the initial physiological state of the cells. The length of the lag phase is determined by the value of h_0 at inoculation and the post-inoculation environment. Essentially it has the same four parameters as the Gompertz model.

Buchanan Model(Eqn 6.)

$$N(t) = \begin{cases} N_0 & \text{if } t \leq t_{lag} \\ N_{max} + r_{max} \cdot (t - t_{lag}) & \text{if } t_{lag} < t < t_{max} \\ N_{max} & \text{if } t \geq t_{max} \end{cases} \quad (9)$$

The Buchanan model was developed by Buchanan et al., 1997 as a three-phase linear model which divides bacterial growth curves in to three specific phases with different equations. t_{max} is the time at which the population size reach N_{max} .

2.2 Data

The data used in this work are collected through lab experiments, sourced from ten different publications (Bae et al., 2014; Bernhardt et al., 2018; Galarz et al., 2016; Gill and DeLacy, 1991; Phillips and Griffiths, 1987; Roth and Wheaton, 1962; Silva et al., 2018; Sivonen, 1990; Stannard et al., 1985; Zwietering et al., 1994). The collection contains growth data from 45 species with various culture conditions (different temperature and 18 types of medium) for varying replications and time. Prior to the use of the data for model fitting, negative values in population size were removed (population size cannot be negative in reality). The models were written in log to the base 10, therefore the population value was transformed to log base 10. The data was subdivided according to its species, culture medium, temperature, replicate time and citation.

2.3 Starting Values Calculation

The following starting values of the parameters were calculated:

- N_0 : population size at the start of the data
- N_{max} : highest population size (equation: $\max(\text{diff}(\text{data}\$Log10N)/\text{mean}(\text{diff}(\text{data}\$Time)))$, where data is the used data and Log10N is the log10 population value)
- r_{max} : maximum growth rate by finding the highest slope of the curve
- t_{lag} : time at which the difference of the log10 population value was at its peak (equation: $\text{data}\$Time[\text{which.max}(\text{diff}(\text{diff}(\text{data}\$Log10N)))]$)

2.4 Model Selection Methods

Three commonly used model selection methods to identify the model best supported by the data (Table 1). The adjusted residual sum of squares (RSS), which regarded as the criteria for the fit quantity of models, are used to selected the data that is suitable for model selection. If the RSS of more than four models is "NA" or less than 0.75, the subset of data will be removed

Table 1: Used commonly model selection methods

Model selection methods	Calculation	References
Adjusted R^2 (RSS)	$R_{adj}^2 = 1 - \frac{RSS(n-p-1)}{\sum (y_i - \bar{y})^2} \frac{n-1}{n}$	Rohlf and Sokal, 1981
Akaike information criterion (AIC)	$AIC = -2\ln[L(\hat{\theta}_p y)] + 2p$	Burnham and Anderson, 2002
Bayesian information criterion (BIC)	$BIC = -2\ln[L(\hat{\theta}_p y)] + p \cdot \ln(n)$	Schwarz et al., 1978

n , sample size; p , count of free parameters; y : data;
 $\ln[L(\hat{\theta}_p|y)]$: likelihood of the model parameters (more precisely, maximum likelihood estimates).
 $\ln[L(\hat{\theta}_p|y)$ given the data, y .

as it may be less informative for comparison than the others. For the data subset of interest, candidate models are compared by evaluating the relative support in the data subset of interest. The level of support for each candidate model was evaluated in following scheme: after models were fitted within each research, the minimal Akaike information criterion (AIC) was determined, and the delta AIC was calculated. The model(s) with the minimal AIC and the model(s) with delta AIC less than 2 was considered as the best model(s) (Burnham and Anderson, 2004). The percentage of the number of times each model successfully converged out of all subsets is calculated as convergence score. The percentage of the number of which a candidate model is the best model out of all of the subset and that of a candidate model is the best out of the subset is collected as fitting score. The score of the Bayesian information criterion (BIC, also known as the Schwarz criterion, or SC) is collected in the same procedure. The model with the higher fitting score is considered to be better supported.

2.5 Computing Languages

Data preparation was carried out in python 3.7.4. Model fitting and comparison was done in R 3.6.1. R was used with packages "minpack.lm", "ggplot", "dplyr" and "tidyr". Python was used with package "pandas", "scipy" and "numpy".

3 Results

17 subset is removed according to RSS value. All of the models visually gave reasonably good fits of the data except for the Buchanan model (Figure 2 for example). Overall, the Logistic model has the highest convergence score and the Gompertz model has the highest fitting score (Figure 1). The fitting scores of logistic model with four parameters was higher than the one

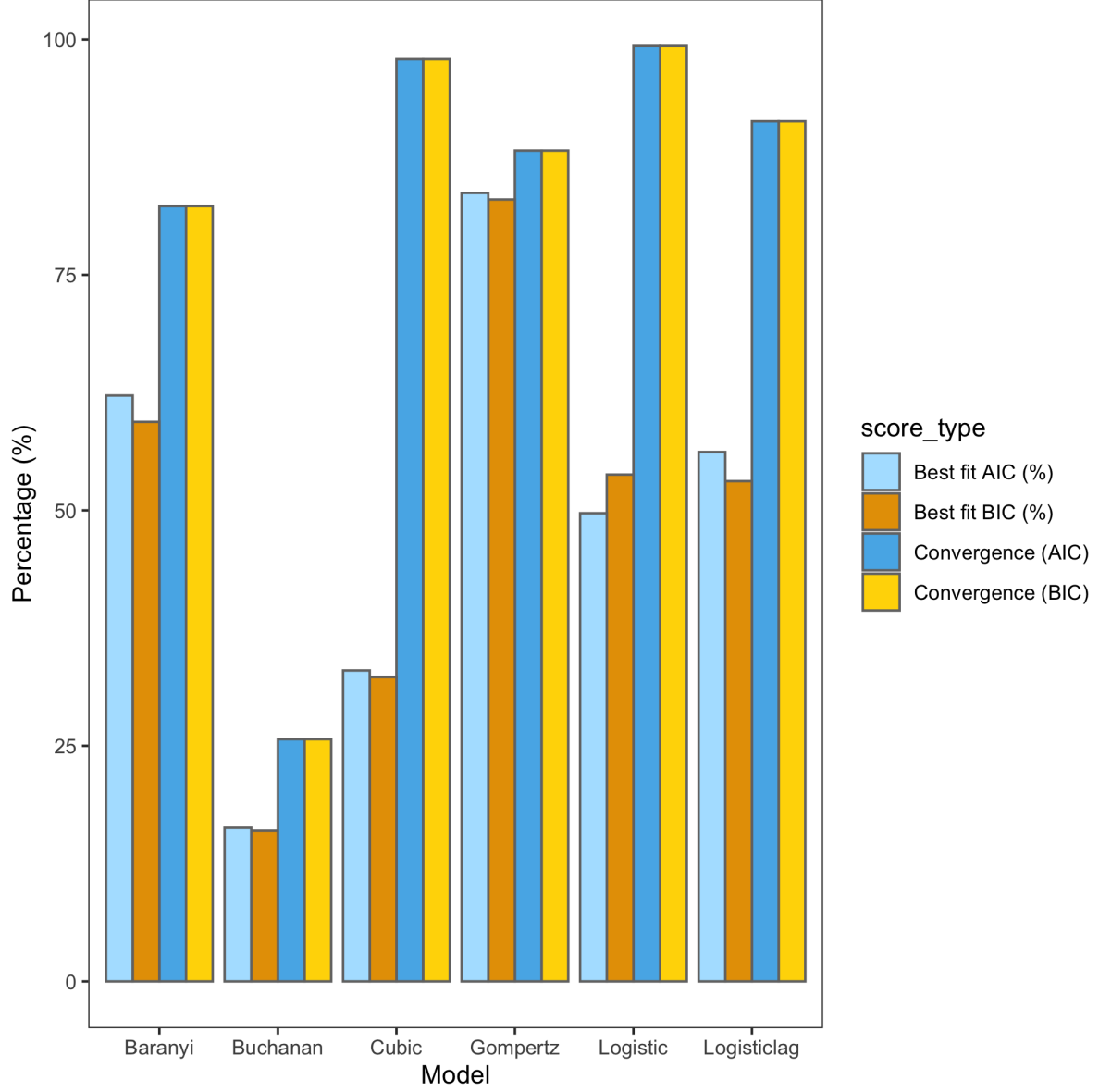


Figure 1: Overall convergence and fitting score of candidate models

of logistic model with three parameters.

4 Discussion

In this work, 5 models were compared statistically and the results showed that for the data used, the Gompertz model can be regarded as the most sufficient models among the candidates to best describe about 83% of the subsets (Table 2). The AIC and BIC fitting scores gave a similar results overall. The BIC fitting score (53.8%) is higher than the AIC fitting score (53.8%) for the three-parameter logistic model and for the four-parameter logistic model is true (BIC: 53.1%,

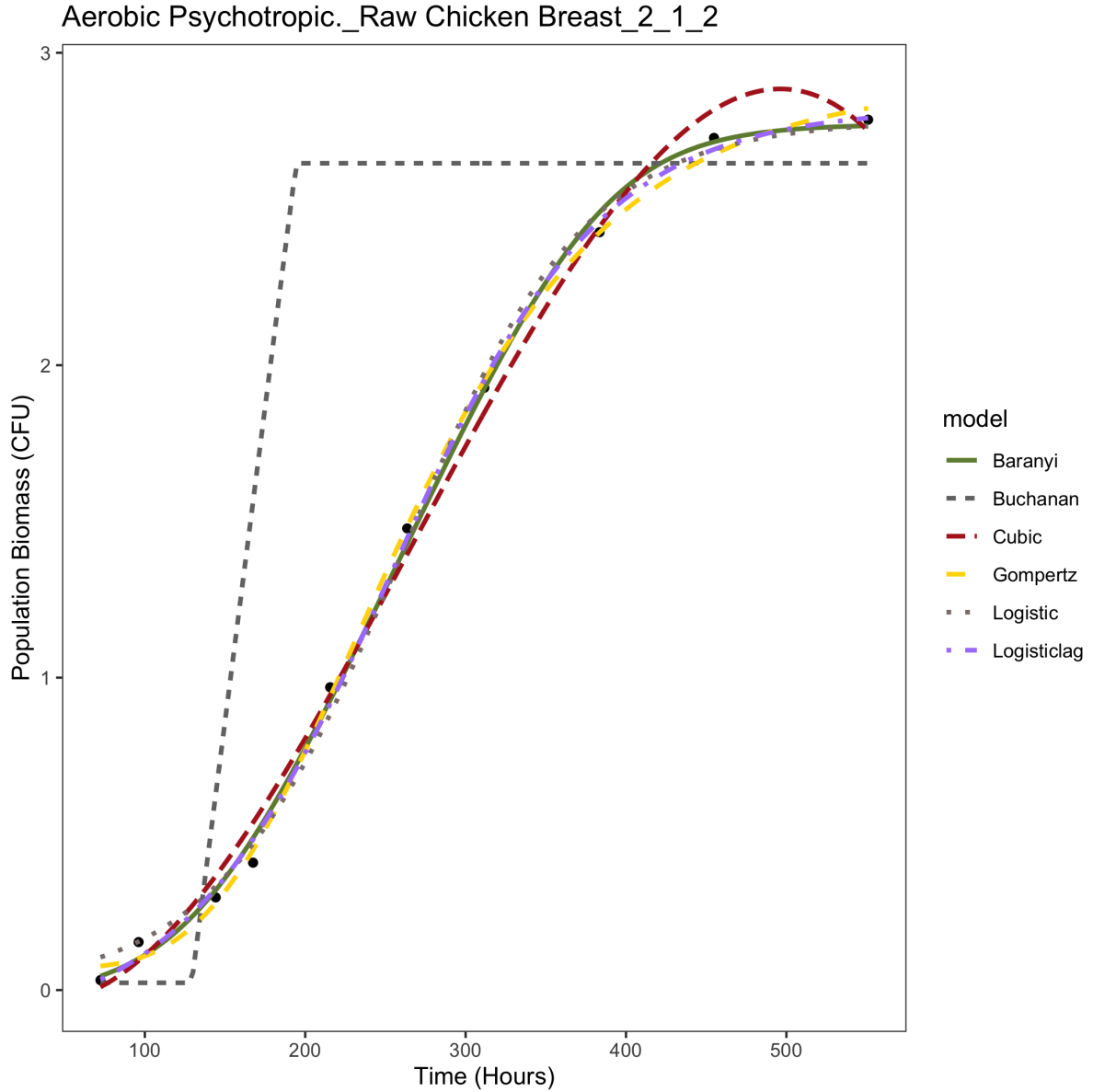


Figure 2: Growth curves of aerobic psychrotrophic bacteria in raw chicken breast at 2 °C fitted with 6 candidate models

98 AIC:56.2%). This is reasonable as the penalty term for the number of parameters in the model
 99 introduced in BIC than in AIC. According to both AIC and BIC fitting, the three-parameter
 100 logistic model and the four-parameter logistic model performed almost the same.
 101 Among all candidate models, the Buchanan model fitted the least number of subsets and re-
 102 ceived lowest fitting scores. The Buchanan model has been developed to emphasise the lag
 103 phase to help reconcile the known bacteria behaviour. However, the microbial data analysed in
 104 this work comes from various experiments which consist of different species grown in different

Table 2: AIC and BIC fitting scores of candidate modesl

Model	Fitted AIC (%)	AIC fitting score (%)	Fitted BIC (%)	BIC fitting score (%)
Cubic	97.9	33.0	97.9	32.3
Logistic	99.3	49.7	99.3	53.8
Logisticlag	91.3	56.2	91.3	53.1
Gompertz	88.2	83.7	88.2	83.0
Baranyi	82.3	62.2	82.3	59.4
Buchanan	25.7	16.3	25.7	16.0

conditions. The growth curves of the bacteria in this case does not always show a sigmoidal shape, which could gave difficulties fitting with the Buchanan model. The fact that the definition of lag is independent from the shape of the growth curve allows model with t_{lag} parameter to model growth without a lag period, which could be the reason of the better performance of the Gompertz and the Baranyi model.

References

- 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.
- József Baranyi and Terry A Roberts. Mathematics of predictive food microbiology. *International journal of food microbiology*, 26(2):199–218, 1995.
- Joey R Bernhardt, Jennifer M Sunday, and Mary I O’Connor. Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity. *The American Naturalist*, 192(6):687–697, 2018.
- Benjamin M Bolker. *Ecological models and data in R*. Princeton University Press, 2008.
- RL Buchanan, RC Whiting, and WC Damert. When is simple good enough: a comparison of the gompertz, baranyi, and three-phase linear models for fitting bacterial growth curves. *Food microbiology*, 14(4):313–326, 1997.
- Kenneth P Burnham and David R Anderson. A practical information-theoretic approach. *Model selection and multimodel inference, 2nd ed.* Springer, New York, 2002.

- Kenneth P Burnham and David R Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304, 2004.
- Liane Aldrighi Galarz, Gustavo Graciano Fonseca, and Carlos Prentice. Predicting bacterial growth in raw, salted, and cooked chicken breast fillets during storage. *Food Science and Technology International*, 22(6):461–474, 2016.
- Angela M Gibson, N Bratchell, and TA Roberts. Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by ph, sodium chloride and storage temperature. *International journal of food microbiology*, 6(2):155–178, 1988.
- CO Gill and KM DeLacy. Growth of escherichia coli and salmonella typhimurium on high-ph beef packed under vacuum or carbon dioxide. *International journal of food microbiology*, 13(1):21–30, 1991.
- Ray Hilborn and Marc Mangel. *The ecological detective: confronting models with data*. Princeton University Press, 1997.
- Jerald B Johnson and Kristian S Omland. Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2):101–108, 2004.
- Robin C McKellar and Xuewen Lu. Primary models. *Modeling microbial responses in food*, pages 21–62, 2004.
- JD Phillips and MW Griffiths. The relation between temperature and growth of bacteria in dairy products. *Food Microbiology*, 4(2):173–185, 1987.
- WE Ricker. Growth rates and models. *Fish physiology*, pages 673–743, 1979.
- F James Rohlf and Robert R Sokal. *Biometry: the principles and practice of statistics in biological research*. Freeman New York, 1981.
- Norman G Roth and Robert B Wheaton. Continuity of psychrophilic and mesophilic growth characteristics in the genus arthrobacter. *Journal of bacteriology*, 83(3):551–555, 1962.
- Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

- 151 Ana Paula Rosa da Silva, Daniel Angelo Longhi, Francieli Dalcanton, and Gláucia Maria
 152 Falcão de Aragão. Modelling the growth of lactic acid bacteria at different temperatures.
 153 *Brazilian Archives of Biology and Technology*, 61, 2018.
- 154 Kaarina Sivonen. Effects of light, temperature, nitrate, orthophosphate, and bacteria on growth
 155 of and hepatotoxin production by *oscillatoria agardhii* strains. *Appl. Environ. Microbiol.*, 56
 156 (9):2658–2666, 1990.
- 157 CJ Stannard, AP Williams, and PA Gibbs. Temperature/growth relationships for psy-
 158 chrotrophic food-spoilage bacteria. *Food Microbiology*, 2(2):115–122, 1985.
- 159 MH Zwietering, Il Jongenburger, FM Rombouts, and KJAEM Van’t Riet. Modeling of the
 160 bacterial growth curve. *Appl. Environ. Microbiol.*, 56(6):1875–1881, 1990.
- 161 MH Zwietering, JC De Wit, HGAM Cuppers, and K Van’t Riet. Modeling of bacterial growth
 162 with shifts in temperature. *Appl. Environ. Microbiol.*, 60(1):204–213, 1994.