

CMEE MiniProject:

Comparison between Mechanistic Model and Phenomenological Model fitting on growth rate data

Student: Hongye Wang (*hw2419@ic.ac.uk*)

Affiliation: Department of Life Sciences. Imperial College London

Word Count: 3603

Contents

1 Abstract	1
2 Introduction	1
2.1 Bacterial growth curve	2
2.2 Models	3
3 Methods	5
3.1 Data Preparation	5
3.2 Find the starting value	5
3.3 NLLS fitting	9
3.4 Model Selection	10
3.5 Computing languages	12
4 Results	12
5 Discussion	17

1 Abstract

In order to estimate bacterial growth kinetics, many sigmoidal functions were compared to fit the growth curves. In this report, the tested models are Classical mechanistic model, Gompertz model, Baranyi model, Buchanan model and Cubic model. The aim of my work is to find the best fitted one. They were fitted by the methods of NLLS fitting. The outcomes were compared and analyzed under the condition of AIC , AIC_c , ΔAIC , ΔAIC_c and Akaike Weight. Eventually, Classical mechanistic model provided the best fit result among all models.

2 Introduction

The use of primary mathematical models with curve fitting software is more and more welcomed in the microbiology field. Food microbiologists use the way of predictive modeling to analyze the bacterial growth data and estimate the microbial safety or shelf life of products(Zwietering et al. 1990). The prediction model mainly expresses the functional relationship between the time and the population of microorganisms, i.e. the response of the microorganisms, and the equation is expressed by a series of specific parameters. The parameters are divided into two types: direct response parameters and indirect

17 response parameters. In recent years, researchers have proposed many mathematical
18 equations describing the dynamic growth of microorganisms, including Gompertz equa-
19 tion, Logistic equation, Baranyi equation and Buchanan equation, which will be described
20 in detail below.

21 The aim of my work is to find the best fitted model which can be used to estimate bacteria
22 growth.

23 **2.1 Bacterial growth curve**

24 The growth curve is represented by the dotted line on the graph of the growth phe-
25 nomenon. Under the proper growth conditions, the bacterial population increased in
26 the schizont way and show an exponential explosion(Eagon 1962). The growth curve
27 drawn with the logarithm of the number of bacterial cells as the ordinate and the growth
28 time as the abscissa, also called a sigmoid curve, which can be divided into four parts,
29 reflecting the four main stages of bacterial growth(Baty & Delignette-Muller 2004): lag
30 phase, logarithmic phase, stationary phase and decline phase. 1. lag phase: generally,
31 bacteria need a brief period to adapt the new environment. The curve in this period is flat
32 and stable, because the bacteria reproduce very little, and the reproduction speed grad-
33 ually increases from zero(Bertrand 2019). We use N_0 to represent the initial value of the
34 bacterial population, and t_{lag} to represent the period of time when the proliferation rate
35 changes from 0 to the maximum rate. 2. logarithmic phase: also called the exponential
36 growth phase. After the preparation period of the lag phase, the bacteria grew extremely
37 fast with a stable geometric progression, and therefore achieved the maximum growth
38 rate(r_{max}), which showed a rapid rise in the growth curve and an exponential increase
39 in the population of bacteria. In the end of logarithmic phase, the maximum population
40 number(N_{max}) should be reached. 3. stationary phase: the growth curve in this period
41 became to a horizontal line, which is call an asymptote(A), but the bacterial population
42 vigor changed greatly. Due to the adverse effects of nutrient consumption in the culture
43 medium, toxic products (organic acids, H_2O_2 , etc.) and the change of pH, the bacterial
44 reproduction rate gradually decreased, and the relative number of bacterial deaths began
45 to increase gradually(Bridges et al. 2001). As a result, the number of bacterial prolifer-

46 ation and death finally balanced. 4. decline phase: with the development of the stable
 47 period, the bacteria multiply slowly and the number of dead bacteria increases signifi-
 48 cantly. The number of viable bacteria was inversely related to the culture time, and the
 49 growth curve began to fall(Novick 1955). However, in this article, we only consider the
 50 first three phases of the bacterial growth curve during the mode fitting process.

51 **2.2 Models**

52 **1. Classical mechanistic model**

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

53 A classical mechanistic model is the logistic equation. Here N_t is population size at time
 54 t , N_0 is initial population size, r is maximum growth rate (AKA r_{max}), and N_{max} is carrying
 55 capacity (commonly denoted by K in the ecological literature).

56 **2. Gompertz model (Zwietering et al. 1990)**

$$N_t = A e^{-e^{\frac{r_{max} e^{t_{lag}}}{A} - t}} + 1 \quad (2)$$

57 Gompertz model is one of the most widely used mathematical model, which has been
 58 modified now to model bacterial growth(Gibson et al. 1988). Here maximum growth rate
 59 (r_{max}) is the tangent to the inflection point, t_{lag} is the x-axis intercept to this tangent (dura-
 60 tion of the delay before the population starts growing exponentially) and A is the asymp-
 61 tote ($A = \ln(N_{max}/N_0)$), N_0 is initial cell culture (Population) density, N_{max} is maximum
 62 population density.

63 **3. Baranyi model(Baranyi et al. 1993)**

$$N_t = N_0 + r_{max} A_t - \ln\left(1 + \frac{e^{r_{max} A_t - 1}}{e^{N_{max} - N_0}}\right) \quad (3)$$

64 **Where:**

$$A_t = t + \frac{1}{r_{max}} \cdot \ln\left(\frac{e^{-r_{max} t + h_0}}{1 + h_0}\right) \quad (4)$$

Here h_0 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. Thus the definition of lag is independent from the shape of the growth curve, and the effect of the previous environment is separated from the effects of the present environment. In this model, r_{max} and h_0 can be related to obtain the t_{lag} :

$$t_{lag} = \frac{\ln(1 + \frac{1}{h_0})}{r_{max}} \quad (5)$$

4. Buchanan model Buchanan model is a three-phase linear model, which use three line segments represent the three phases of bacterial growth(Buchanan et al. 1997).

$$Lag - phase : N_t = N_0 \quad t \leq t_{lag} \quad (6)$$

$$Logarithmic - Phase : N_t = N_0 + \mu(t - t_{lag}) \quad t_{lag} < t < t_{max} \quad (7)$$

$$Stationary - Phase : N_t = N_{max} \quad t \geq t_{max} \quad (8)$$

where: N_t is log of the population density at time t ; N_0 is log of the initial population density; N_{max} is log of the maximum population density supported by the environment; t is Elapsed time; t_{lag} is the time when the lag phase ends; t_{max} is the time when the maximum population density is reached; μ represents specific growth rate(Buchanan & Cygnaarowicz 1990).

5. cubic polynomial model

$$N_t = B_0 + B_1t + B_2t^2 + B_3t^3 \quad (9)$$

This is a phenomenological model, with the parameters B_0 , B_1 , B_2 and B_3 lacking any mechanistic interpretation. t is the independent variable

3 Methods

3.1 Data Preparation

The first thing we need to do is to browse and reorganize the data we used for model fitting (Fig.1).

	X	Time	PopBio	Temp	Time_units	PopBio_units	Species	Medium	Rep	Citation
0	1	669.879518	0.283276	5	Hours	OD_595	Chryseobacterium.balustinum	TSB	1	Bae, Y.M., Zheng, L., Hyun, J.E., Jung, K.S., ...
1	2	646.987952	0.283342	5	Hours	OD_595	Chryseobacterium.balustinum	TSB	1	Bae, Y.M., Zheng, L., Hyun, J.E., Jung, K.S., ...
2	3	622.891566	0.285151	5	Hours	OD_595	Chryseobacterium.balustinum	TSB	1	Bae, Y.M., Zheng, L., Hyun, J.E., Jung, K.S., ...
3	4	597.590361	0.281746	5	Hours	OD_595	Chryseobacterium.balustinum	TSB	1	Bae, Y.M., Zheng, L., Hyun, J.E., Jung, K.S., ...
4	5	574.698795	0.273117	5	Hours	OD_595	Chryseobacterium.balustinum	TSB	1	Bae, Y.M., Zheng, L., Hyun, J.E., Jung, K.S., ...

Figure 1: Initial database with the column headers

I categorize the data by species, citation and rep, and create unique IDs for each group in the data frame, which is more convenient for building subsets when modeling fitting and plotting. Then, remove unusual data, such as those with a time of less than zero, or an empty set. After that, filter out the data subsets with less than 5 data points, because 5 is the minimum number of data points needed to fit the models. At last, log the data in the 'PopBio' column, store the results in a new column and rename them as 'log-popbio'. After preparation work, there are 295 available subsets grouped by ID, each group will be used to fit five models.

3.2 Find the starting value

It is important to finding appropriate starting value for each parameter in Non-linear Least Squares(NLLS) fitting method. Table.1 shows how many parameters each model has and what they are.

Model	Parameters
Classical mechanistic model	N_0, N_{\max}, r_{\max}
Gompertz model	$A, t_{\text{lag}}, r_{\max}$
Baranyi model	$N_0, N_{\max}, r_{\max}, h_0$
Buchanan model	$N_0, N_{\max}, t_{\text{lag}}, t_{\max}, \mu$
cubic polynomial model	B_0, B_2, B_3, B_4

Table 1: Five selected models and their parameters

96 Since all the models have been re-parameterized to substitute the matical parameters
97 with A, t_{lag} and r_{\max} (except the cubic) which made them can be calculated directly, now
98 the models are easily to share the starting values in terms of same parameters. The
99 method of finding starting value is as follows:

100 N_0 : the initial population size of each ID group.

101 N_{\max} : the maximum population size of each ID group.

102 r_{\max} : the steepest slope of the growth curve. Searching for the maximum slope is little
103 more complicated. I sorted the points in the subset by time order, then draw a straight line
104 every four points using Ordinary Least Squares(OLS) starting from the first point. Finally,
105 comparing these line, the maximum slope is the value of r_{\max} (Fig.2).

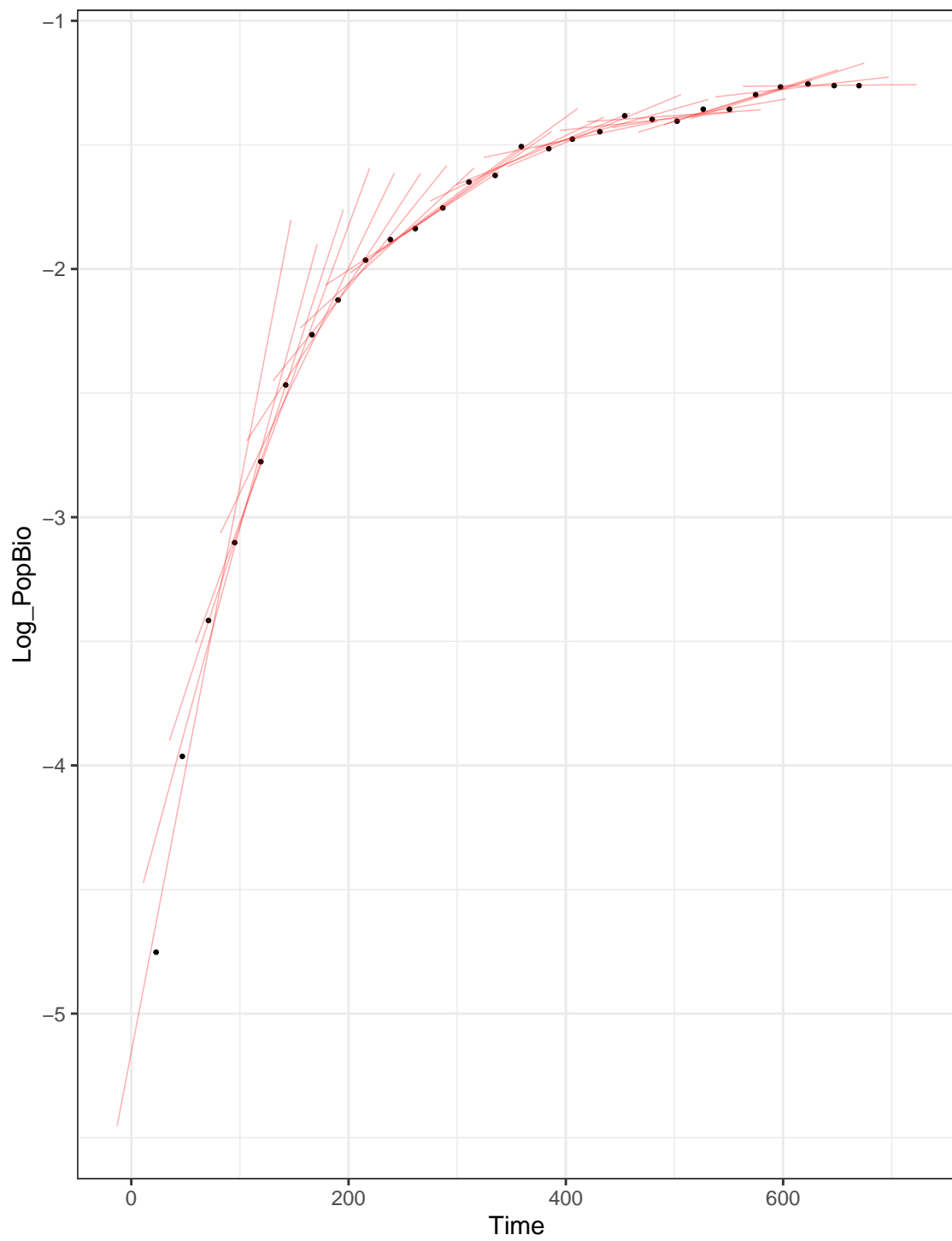


Figure 2: Plotting rolling regression in a data subset (ID is 1)

¹⁰⁶ lag: The x-intercept created by the line with maximum slope.(Fig.3)

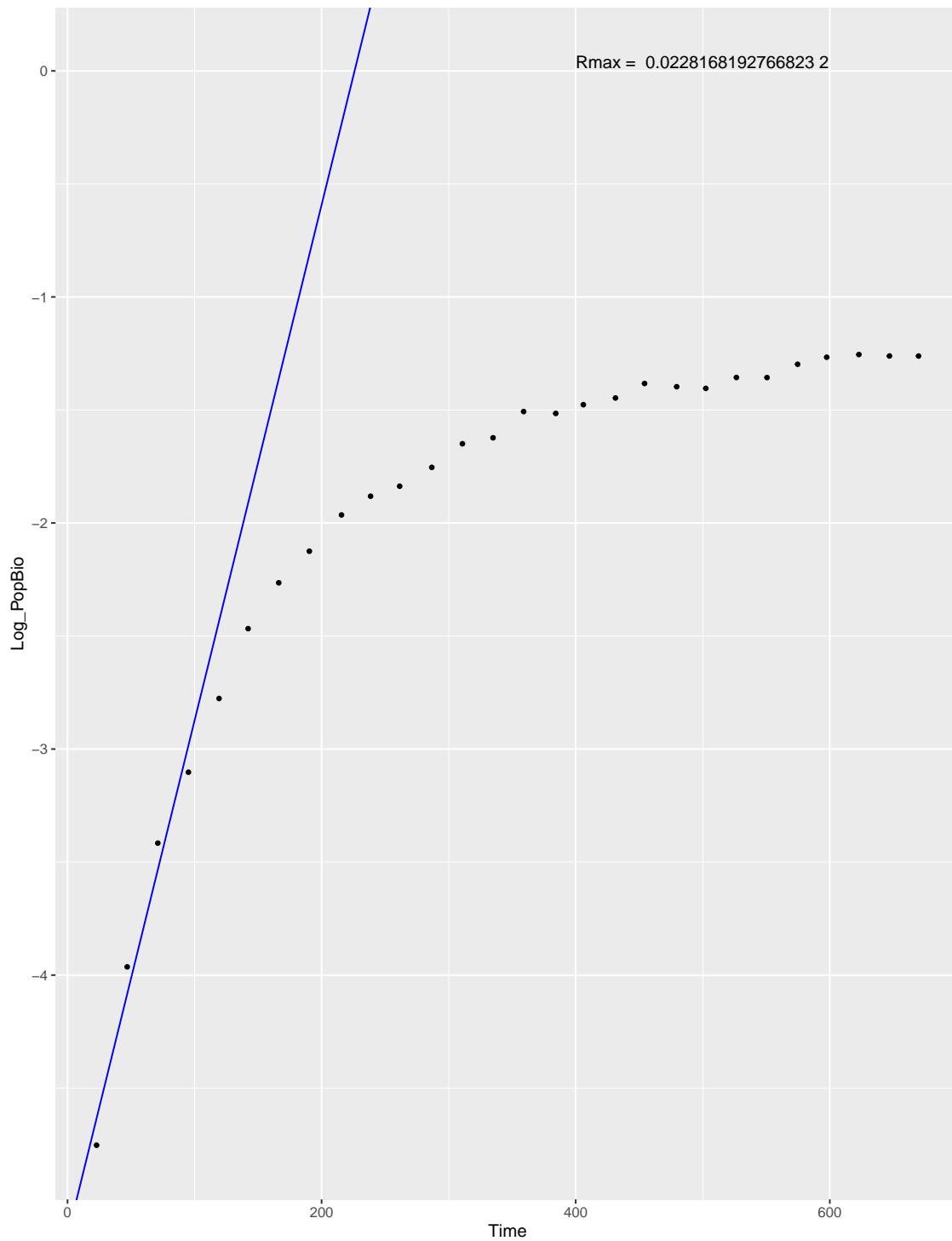


Figure 3: The line with the maximum slope and show the R_{\max} value on the top

107 t_{\max} : The time when population size reached the maximum value in each group.

108 A : When the growth curve is defined as the logarithm of the number of bacteria plotted
 109 against time, A is the asymptote of the growth curve, which equals to $\ln (N_{\max}/N_0)$ (Fig.

110 4).

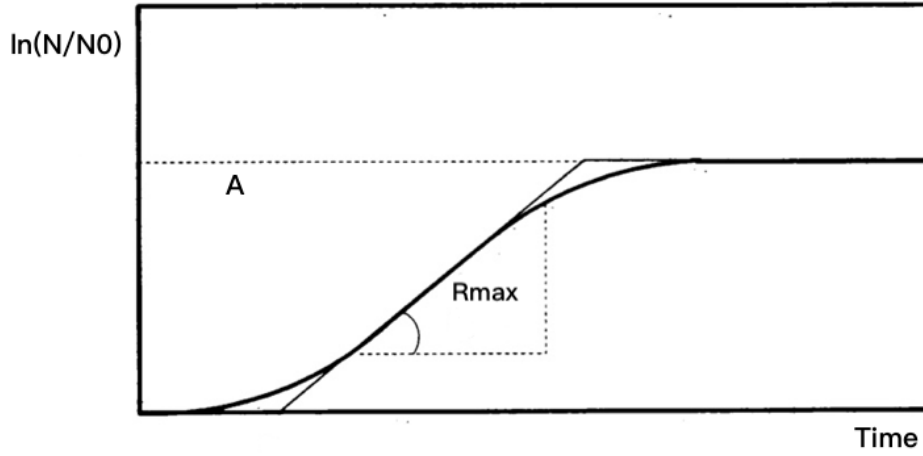


Figure 4: A growth curve

111 H_0 : In baranyi model, h_0 can be related to t_{lag} and r_{max} , which made it easily to be
112 calculated:

$$t_{lag} = \frac{1 + \frac{1}{h_0}}{r_{max}} \quad (10)$$

113 So:

$$h_0 = \frac{1}{e^{r_{max}t_{lag}} - 1} \quad (11)$$

114 μ : In Buchanan model, μ value is calculated by the equation below(Damert 1994):

$$\mu = \frac{N_{max} - N_0}{t_{max} - t_{lag}} \quad (12)$$

115 After all the starting values are confirmed, I built new columns to store them in data frame

116 3.3 NLLS fitting

117 Here, I use the Python package LMFIT(Newville et al. 2016) for each model to apply
118 model fitting on each curve by NLLS method. The main task for doing a non-linear least-
119 squares fit of a model to data is to write an objective function that takes the values of

the fitting variables and calculates either a scalar value to be minimized, typically in the least-squares sense. Also, the objective function should return the value to be minimized. So, what I need to do is calculating objective residual to be minimized from parameters. Parameter is the quantity to be optimized in all minimization problems, the parameters are given the starting value as their initial value when first try. Then, assign parameters randomly from the range of normal distribution with the initial value as the axis. It is hope that the LMFIT can find the best parameter values for each model within the maximum try times. The residual is calculated as (model – data) with the best give parameters. Note that the cubic model is a polynomial linear model and the best way to fit this model might be the Ployfit funtion in Numpy for Python, but in this study I still used LMFIT package because I want to test if this package designed for non-linear model fitting could also applied in linear model fitting. It turns out that LMFIT not only can fit the non-linear model, but also available in linear model.

3.4 Model Selection

Many model selection methods are used to find the optimal model, like R^2 , AIC , BIC . Here I choose the Akaike information criterion (AIC) as my main method. the AIC value of the model is the following:

$$AIC = 2\rho + \ln(L) \quad (13)$$

Where ρ is the number of estimated parameters in the model; L is the maximum value of the likelihood function for the model(McElreath 2016).

When the sample size is small, there is a substantial probability that AIC will select models that have too many parameters, i.e. that AIC will over fit(McQuarrie & Tsai 1998). To address such potential over fitting, AIC_c was developed: AIC_c is AIC with a correction for small sample sizes(Cavanaugh 1997).

$$AIC_c = AIC + \frac{2\rho^2 + 2\rho}{n - \rho - 1} \quad (14)$$

Comparing to the other methods, the formula for the Bayesian information criterion (BIC)

144 is similar to AIC , but with a different penalty for the number of parameters. With AIC the
 145 penalty is 2ρ , whereas with BIC the penalty is $\ln(n)\cdot\rho$.

146 A comparison of AIC/AIC_c and BIC is given by Burnham and Anderson (Burnham
 147 & Anderson 2002), with follow-up remarks by Burnham and Anderson (Burnham & An-
 148 derson 2004). The authors show that AIC/AIC_c can be derived in the same Bayesian
 149 framework as BIC , just by using different prior probabilities. In BIC , though, each can-
 150 didate model has a prior probability of $1/R$, where R is the number of candidate models.
 151 Such a derivation is 'not sensible', because the prior should be a decreasing function of
 152 ρ . Additionally, the authors present a few simulation studies that suggest AIC_c tends to
 153 have practical/performance advantages over BIC (Burnham & Anderson 2004).

154 As for R^2 , although it is the simplest way to compare two models in terms of fitting (eq15).
 155 It calculated from Residual Sum of Squares (RSS) and Total Sum of Squares (TSS).
 156 However, it neglects the complexity of the model, which would lead to a situation that a
 157 very complicated model with lots of parameters has been chosen just because it con-
 158 verges best. In order to avoid this problem, AIC provide a better choice because it
 159 considers about the complexity of models and penalises the over-fitting (Bozdogan 1987).
 160 Equation of calculating AIC contains the number of parameters in model (eq16).

$$161 \quad R^2 = 1 - \frac{RSS}{TSS} \quad (15) \quad AIC = N \ln \frac{RSS}{N} + 2\rho \quad (16)$$

162 After getting the AIC and AIC_c for each curve, ΔAIC and ΔAIC_c would then be cal-
 163 culated as the difference between the lowest $AIC(AIC_c)$ and the $AIC(AIC_c)$ for each
 164 model. If ΔAIC or ΔAIC_c was less than or equal to 2, then the corresponding model can
 165 be regarded as the best model (Burnham & Anderson 2004).

166 The Akaike weight $W_i(AIC)$ was then calculated for each model by below equation. This
 167 method can provide the likelihood of each model being the best choice, which promotes
 168 the interpretation of comparing models. $W_i(AIC_c)$ had also been calculated for analy-
 169 sis (Wagenmakers & Farrell 2004).

$$W_i = \frac{\exp\{-\frac{1}{2}\Delta_i\}}{\sum_{j=1}^R \exp\{-\frac{1}{2}\Delta_j\}} \quad (17)$$

170 **3.5 Computing languages**

171 Python 3.5.2 was used for arranging data, estimating starting values of parameters in
172 population growth models and model fitting with NLLS. Using library pandas to manipu-
173 late the large database easily. It is quickly to use Numpy and Scipy to get the estimate
174 values(McKinney et al. 2011). Choosing Python's LMFIT package rather than R is be-
175 cause when facing the large database and complicated calculating process, Python often
176 does better and faster than R.

177 R version: 3.2.3. R was used in model selection stage by its' calculating function because
178 it is user-friendly. The drawing process also used R, because the ggplot2 library can
179 produce very accurate and high quality figures(Ginestet 2011).

180 GNU bash, version 4.3.48(1). The Bash Shell was used to tie all the scripts in this project
181 as a workflow and compile the report from latex script to pdf file.

182 **4 Results**

183 After manipulating original database, there are 295 curves. Among the five models, ex-
184 cept the classical mechanistic model has several unfitted subsets, the success fitting rate
185 of other models are 100%. So, I plot the models in one data subset to observe the actual
186 fitting imagine.

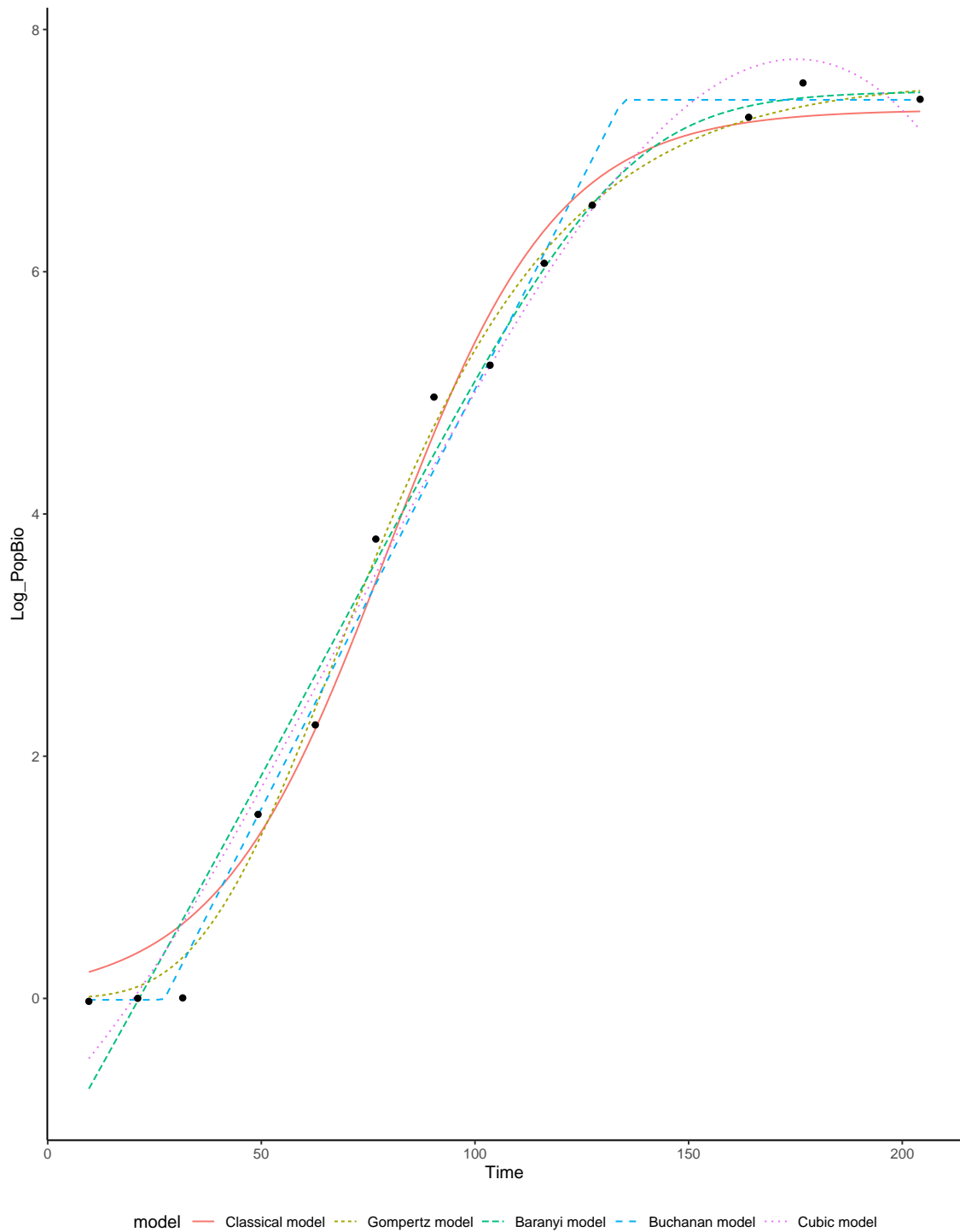


Figure 5: Comparison five models' curves. X axis represents the time and Y axis represents logged population size of *Pseudomonas*. The data was taken from (Galarz, L.A., Fonseca, G.G. and Prentice, C., 2016) The analysis information for comparison between each model are listed in the table below.

Model	R^2	AIC	AIC_c	ΔAIC	ΔAIC_c	Weight(AIC)	Weight(AIC_c)
Classical	0.989	-24.93	-22.26	13.32	13.32	0.0013	0.0013
Gompertz	0.996	-38.25	-35.58	0	0	0.9922	0.9983
Baranyi	0.986	-20.28	-15.28	17.97	20.31	0.0001	$3.88e - 5$
Buchanan	0.992	-28.12	-19.55	10.13	16.03	0.0062	0.0003
Cubic	0.987	-20.89	-15.89	17.36	19.70	0.0002	$5.28e - 5$

Table 2: R^2 , AIC , AIC_c , ΔAIC , ΔAIC_c and Akaike Weight for each model curve in fig2

187 It seems that all the five models are fitted nicely in this data subset, so it's hard to say
188 which one is the best from the eyeballing way. However, from the details in table2, we can
189 clearly find that the Gompertz model is the best fitted model for this data subset, because
190 it has the minimum value of AIC , AIC_c , ΔAIC and ΔAIC_c , also with the maximum
191 value in Weight(AIC) and Weight(AIC_c).

192 Let's have an overall statistical analysis among the five models. Tables below show the
193 statistical analysis information in overall scale.

Model	$\Delta AIC \leq 2$	$\Delta AIC_c \leq 2$
Classical	139	169
Gompertz	68	79
Baranyi	91	73
Buchanan	82	45
Cubic	133	65

Table 3: ΔAIC and ΔAIC_c in each model less than 2 will be counted across all curves

194 From table3, it is observed that Classical mechanistic model has the largest number of
195 data subset with both ΔAIC and ΔAIC_c less than 2, that means the Classical mecha-
196 nistic model has the best fitting curve. Note that Cubic model also has a high proportion
197 in the number of ΔAIC less than 2, but the number of ΔAIC_c less than 2 is decreased
198 dramatically. So I made two box plotting to compare the AIC_c values and Weight(AIC_c)
199 values of each model.

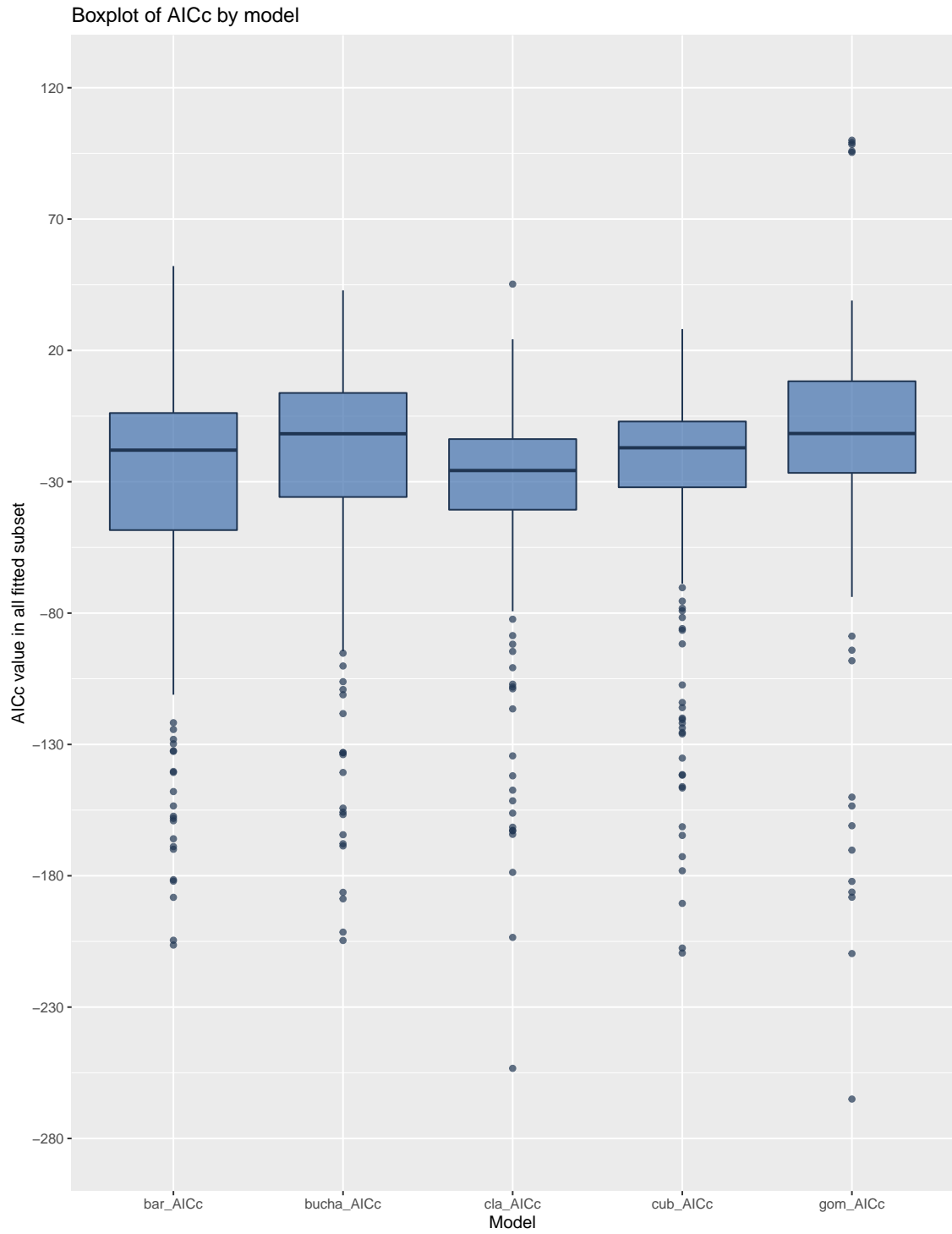


Figure 6: Box plot, shows the AIC_c value distribution in each model

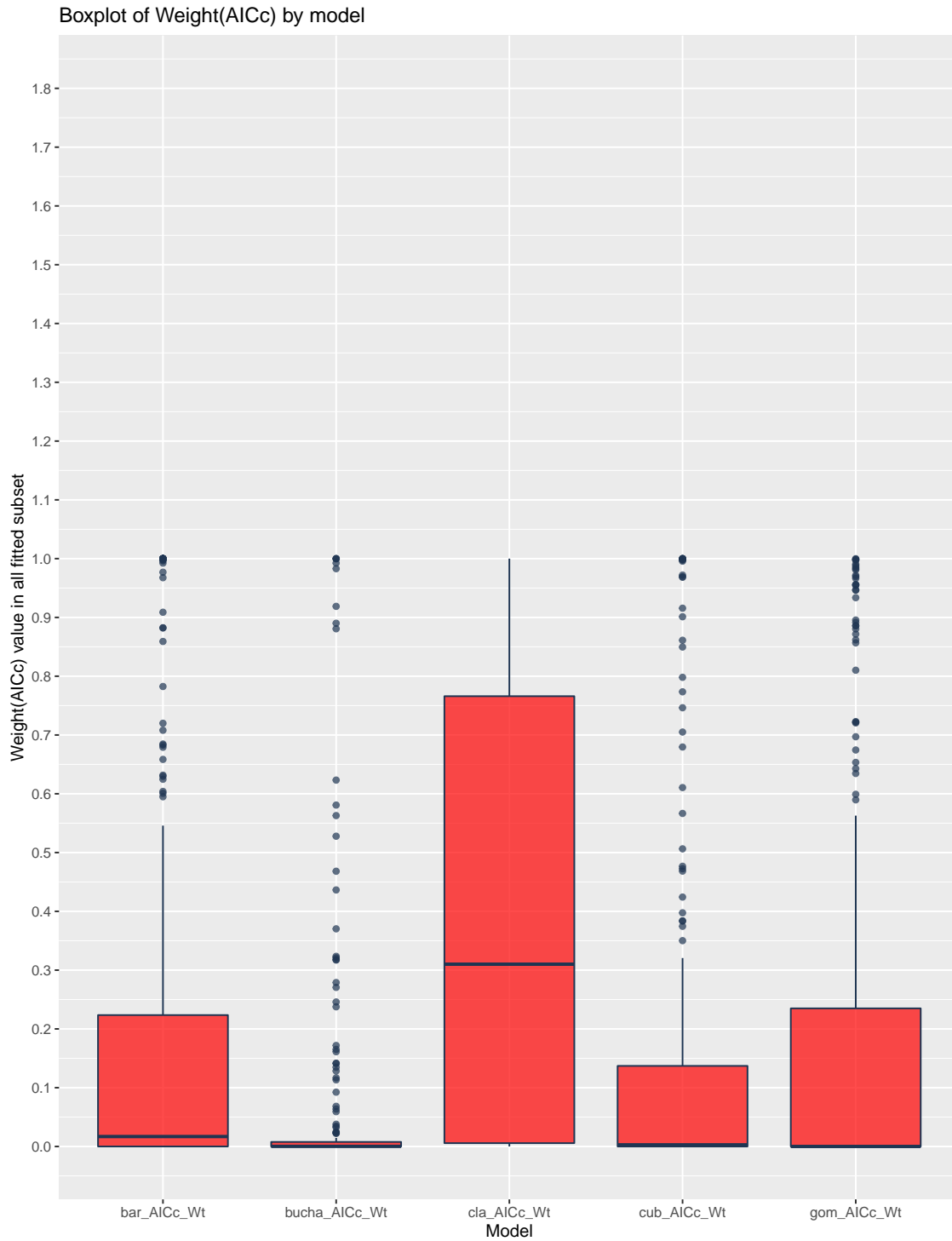


Figure 7: Box plot, shows the Akaike Weight (AICc) distribution in each model

200 In Fig6, the classic mechanistic model has a slight advantage over other models in the
 201 average value of $AICc$. In Fig7, this advantage significantly increased in the Weight
 202 ($AICc$) scale. Thus proves that the classic mechanistic model is the best model for this

203 database. Also, there is a table shows all the mean statistic information of the models.

Model	Mean R^2	Mean AIC	Mean AIC_c	Mean Akaike Weight(AIC)	Mean Akaike Weight(AIC_c)
Classical	0.8681	-31.9921	-34.2791	0.2628	0.3926
Gompertz	-1.0264	-16.3720	-15.1259	0.1343	0.1876
Baranyi	0.8332	-30.7859	-33.8439	0.1807	0.2038
Buchanan	0.8091	-30.17447	-24.9660	0.1769	0.0693
Cubic	0.8988	-31.8501	-30.4140	0.2452	0.1467

Table 4: Mean values of R^2 , AIC , AIC_c , Akaike Weight(AIC) and Akaike Weight(AIC_c) in each model.

204 5 Discussion

205 Although the classical mechanistic model did not match 100% of all subsets successfully,
 206 from the final statistics, it is still the best model to fit the bacteria growth curve under the
 207 given database. The support evidence is shown in table3, table4, the classical mechanis-
 208 tic model owns the minimum value of mean AIC and mean AIC_c , the most times to be
 209 the best model ($\Delta AIC_c \leq 2$) and the biggest likelihood of being the best model (Akaike
 210 weight)(Sakamoto et al. 1986). As for the other models, they have shown good perfor-
 211 mance in the fitting process as well. Like Gompertz model, it used to be the best fit in
 212 some data subsets, however, it is not satisfactory in other data subsets. After my obser-
 213 vation, I think the reason may be related to the size of the population in the data. In some
 214 subsets that use larger orders of magnitude as units, the population size becomes very
 215 small, which makes the values of A and t_{lag} extremely small. In this situation, Gompertz
 216 model is easy to linearize and the outcome become not so well. cubic is a classic mathe-
 217 matical model, and it is also a universal model for many biological phenomena. Although
 218 the success rate of its model fitting is 100%, through the plotting of some subsets, it can
 219 be found that the fitting curve of cubic model is more inconsistent with the actual situation
 220 compared with other models. Because the cubic model does not have any biologically
 221 significance(Brunner et al. 2006), it is not recommended as a predictive model for esti-
 222 mating bacterial growth trends. What surprised me most is the buchanan model. As a
 223 three-segment linear model, I was not optimistic about it when I first saw it. I think the
 224 fitting result may be very poor because it looks too different from the growth curve. But

the final fitting results are very nice, not only all the subsets are fitted successfully but also it has lower AIC_c and higher fitting likelihood than some other models. The reason may belong to its biological similarity with the characteristics of the bacterial growth. Each of the three-line segments correspond to a bacterial growth period. The three-phase linear model seems to be the simplest, effective primary model that can be used readily with curve fitting software to estimate bacterial growth kinetics(Buchanan et al. 1997).

References

- Baranyi, J., Roberts, T. & McClure, P. (1993), 'A non-autonomous differential equation to model bacterial growth', *Food microbiology* **10**(1), 43–59.
- Baty, F. & Delignette-Muller, M.-L. (2004), 'Estimating the bacterial lag time: which model, which precision?', *International journal of food microbiology* **91**(3), 261–277.
- Bertrand, R. L. (2019), 'Lag phase is a dynamic, organized, adaptive, and evolvable period that prepares bacteria for cell division', *Journal of bacteriology* **201**(7), e00697–18.
- Bozdogan, H. (1987), 'Model selection and akaike's information criterion (aic): The general theory and its analytical extensions', *Psychometrika* **52**(3), 345–370.
- Bridges, B. A., Foster, P. L. & Timms, A. R. (2001), 'Effect of endogenous carotenoids on "adaptive" mutation in escherichia coli fc40', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **473**(1), 109–119.
- Brunner, I., Herbst, M., Lerche, W. & Walcher, J. (2006), 'Matrix factorizations and mirror symmetry: The cubic curve', *Journal of High Energy Physics* **2006**(11), 006.
- Buchanan, R. L. & Cygnarowicz, M. L. (1990), 'A mathematical approach toward defining and calculating the duration of the lag phase', *Food Microbiology* **7**(3), 237–240.
- Buchanan, R., Whiting, R. & Damert, W. (1997), '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.
- Burnham, K. P. & Anderson, D. R. (2002), 'Model selection and'.

- 252 Burnham, K. P. & Anderson, D. R. (2004), 'Multimodel inference: understanding aic and
253 bic in model selection', *Sociological methods & research* **33**(2), 261–304.
- 254 Cavanaugh, J. E. (1997), 'Unifying the derivations for the akaike and corrected akaike
255 information criteria', *Statistics & Probability Letters* **33**(2), 201–208.
- 256 Damert, W. (1994), 'Abacus: Interactive program for nonlinear regression analysis',
257 *QCPE Bull* **14**, 61.
- 258 Eagon, R. G. (1962), 'Pseudomonas natriegens, a marine bacterium with a generation
259 time of less than 10 minutes', *Journal of bacteriology* **83**(4), 736–737.
- 260 Gibson, A. M., Bratchell, N. & Roberts, T. (1988), 'Predicting microbial growth: growth
261 responses of salmonellae in a laboratory medium as affected by ph, sodium chloride
262 and storage temperature', *International journal of food microbiology* **6**(2), 155–178.
- 263 Ginestet, C. (2011), 'ggplot2: elegant graphics for data analysis', *Journal of the Royal*
264 *Statistical Society: Series A (Statistics in Society)* **174**(1), 245–246.
- 265 McElreath, R. (2016), *Statistical rethinking: A Bayesian course with examples in R and*
266 *Stan*, Vol. 122, CRC press.
- 267 McKinney, W. et al. (2011), 'pandas: a foundational python library for data analysis and
268 statistics', *Python for High Performance and Scientific Computing* **14**(9).
- 269 McQuarrie, A. D. & Tsai, C.-L. (1998), *Regression and time series model selection*, World
270 Scientific.
- 271 Newville, M., Stensitzki, T., Allen, D. B., Rawlik, M., Ingargiola, A. & Nelson, A. (2016),
272 'Lmfit: Non-linear least-square minimization and curve-fitting for python', *Astrophysics*
273 *Source Code Library*.
- 274 Novick, A. (1955), 'Growth of bacteria', *Annual Reviews in Microbiology* **9**(1), 97–110.
- 275 Sakamoto, Y., Ishiguro, M. & Kitagawa, G. (1986), 'Akaike information criterion statistics',
276 *Dordrecht, The Netherlands: D. Reidel* **81**.
- 277 Wagenmakers, E.-J. & Farrell, S. (2004), 'Aic model selection using akaike weights', *Psy-*
278 *chonomic bulletin & review* **11**(1), 192–196.

279 Zwietering, M., Jongenburger, I., Rombouts, F. & Van't Riet, K. (1990), 'Modeling of the
280 bacterial growth curve', *Appl. Environ. Microbiol.* **56**(6), 1875–1881.