

IMPERIAL COLLEGE LONDON

DEPARTMENT OF LIFE SCIENCE

Time Lag Is a Worthwhile Parameter in Mechanistic Modelling of Bacterial Growth

Author:

Rui Zhang

Supervisor:

Samraat Pawar

Date: 1/21/2021

Word Count: 2381

Abstract

There are a lot of different models fitting population growth which can be applied to the analysis of ecosystem dynamics and emergent functional characteristics. This project focuses on cubic polynomial model, Logistic model and Gompertz model, trying to find out which model fits real microbes population growth data the best. After using R to fit models and calculating Akaike information criterion, Schwarz criterion and the residual sum of squares, Gompertz model shows the best fitting performance among these 3 models. The parameter time lag is worthy to being added into microbial growth model considering both fit goodness and model complexity.

1 Introduction

Population growth equation is necessary in dynamic models to describe the development of the ecosystem and interactions between the population and environment (Gamito 1998). Carrying capacity, population growth rate and metabolic rate of population can all be affected by environment and then results in changes of population dynamics and ecosystem dynamics (Bernhardt et al. 2018). Totally comprehension of population growth model is necessary to estimate extended domains, such as marine fisheries productivity (Sarker & Wiltshire 2017), disease transmission (Aldila & Seno 2019), even the direction of evolution (Hendry & Green 2017).

There are also many different population growth models, involving phenomenological and mechanistic models. Cubic polynomial model is the simplest phenomenological mathematical linear model, with 4 parameters, to describe population growth. For mechanistic models, a classical, somewhat mechanistic model is the logistic equation with 3 parameters (Peleg & Shetty 1997). However this model failed to match lag phase of population growth. To capture the lag phase, more complicated bacterial growth models have been designed, one of which is the modified Gompertz model with 4 parameters (Zwietering et al. 1990), which is the most frequently used model

32 in the research to fit population growth. When comparing these models,
33 a 4 parameters totally phenomenological model, a 3 parameter mechanistic
34 model and a 4 parameters mechanistic model, Zwietering used t test and F
35 test to compare models' statistical sufficient to describe population growth
36 and their ease of use (Zwietering et al. 1990).

37

38 Now there are some more scientific model selection criterion to answer the
39 question whether it is worthwhile to add a new parameter to improve the pre-
40 cise of model fitting. The residual sum of squares, aka R^2 , is a naive method
41 to measure fit. But only considering fit is not enough. The Akaike informa-
42 tion criterion (AIC) (Burnham 2002) calculates the Kullback-Leibler infor-
43 mation lost during approximation full reality with the fitted model, which
44 contains a bias correction factor with respect to model complexity. Schwarz
45 criterion, also named Bayesian information criterion (BIC), is another model
46 selection criterion considering both fit and complexity. However, it includes
47 a penalty term dependent on dataset size and tends to choose simpler mod-
48 els (Johnson & Omland 2004). Generally, AIC is preferred by researchers
49 because of based on Kullback-Leibler information.

50

51 Usually, people favor simpler models due to Ockham's razor (aka parsimony
52 principle) if models describe the same pattern adequately (Peleg &
53 Corradini 2011). To find out whether time lag is a worthwhile parameter to
54 be introduced in mechanistic model fitting and whether a more complicated
55 form of equation works better than totally phenomenological linear model,
56 I fit cubic polynomial, Logistic and Gompertz models to a large amount of
57 microbes population growth experiments data with R. Bacteria cultivation
58 is easy and standardized, and it is convenient and fast to get population
59 growth data. Furthermore, bacteria appears to have a large biodiversity
60 and large habitat diversity which plays an important role in human health,
61 food safety, fermentation engineering, environment and so on. R is a plat-
62 form that is mature and convenient to fit ecological model and analyze data,
63 based on numerous support packages, help documents, simple installation

64 and convenient high-level syntax (Bolker et al. 2013). By means of compar-
65 ing statistical measure of model fitting, we can discover which model is the
66 most suitable to describe microbes population growth, especially bacterial
67 growth.

68 **2 Materials & Methods**

69 **2.1 Materials**

70 The dataset contains 4388 records, collected from 10 lab experiments data
71 across the world, containing the change in biomass or population of microbes
72 over time. The species in these microbes data include bacteria and mobile
73 phytoplankton, which play an important role in food safety and environment.
74 Time range of the dataset is also quite large, varying from 1962 to 2018,
75 which could guarantee the universality of model comparison conclusions.

76 **2.2 Methods**

77 **2.2.1 Computing tools**

78 Data processing, model fitting and results visualization are performed with
79 R due to its excellent graphing capability and numerous packages for statisti-
80 cal analysis and data handling, including reshape2, minpack.lm and ggplot2
81 packages used in this program. Python is an ideal tool to build an au-
82 tomated workflow to analyze data with the subprocess module, especially
83 useful when the program involves many different languages. Therefore, I use
84 shell commands in Python script to run R scripts and compile LATEX to
85 create the report, which guarantees this program is fully reproducible.

86 **2.2.2 Data management**

87 In this program, two main variables of interest are time and population.
88 Therefore, I create a new ID with all of the other variables, with which I
89 divide the total dataset into 285 subdatasets. Then I delete negative time

90 and population biomass, during which I find an unreasonable dataset. This
 91 subdataset contains 29 records but 19 biomass values are negative, so I delete
 92 this dataset totally. Finally there are 284 datasets, containing 4284 records
 93 to fit cubic polynomial model, Logistic model and Gompertz model.

94 **2.2.3 Model fitting**

95 There I choose cubic polynomial model, Logistic model and Gompertz model
 96 to fit all of 284 datasets. Because of subsequent AIC and BIC comparison,
 97 the formation of biomass or population should be the same. To guarantee
 98 all of these equation have the same response variable, I transform cubic
 99 polynomial model and Logistic model into log scale.

100 The equation of cubic polynomial model is:

$$\log(N_t) = \log(a + bt + ct^2 + dt^3) \quad (1)$$

101 The equation of Logistic model is:

$$\log(N_t) = \log\left(\frac{N_0 K e^{rt}}{K + N_0(e^{rt} - 1)}\right) \quad (2)$$

102 The equation of Gompertz model is:

$$\log(N_t) = N_0 + (N_{max} - N_0)e^{-e^{\frac{r_{max}e^{\frac{t-t_{lag}}{(N_{max}-N_0)\log(10)}+1}}}{t}} \quad (3)$$

103 a, b, c, d are all parameters without biological meaning. N_t is biomass or
 104 population value at time t, N_0 is biomass or population value at initial con-
 105 dition, N_{max} and K are carrying capacity, r and r_{max} are maximum growth
 106 rate and t_{lag} is time lag which is the duration of delay before the population
 107 grow exponentially.

108

109 Before model fitting, I transform biomass and population data into log scale
 110 and then fit experiments data to models. For the simple linear phenomeno-
 111 logical model, cubic polynomial model, I use R to fit linear model directly.
 112 For Logistic and Gompertz nonlinear models, I need to calculate and choose

113 start value of parameters. N_0 start value is the minimum of biomass or
 114 population. N_{max} and K are same, whose start value is the maximum of
 115 biomass or population. Start value of r is calculated by linear model, choos-
 116 ing the middle 70 percents of biomass or population value range, drawing
 117 a straight line and setting the value of slope as start value. t_{lag} starting
 118 value is set as the time of the largest second order derivative of population
 119 growth point. Furthermore, I sample start values 400 times to find out the
 120 best start value combination.

121

122 After model fitting, I calculate AIC, BIC and R^2 of these model fittings.
 123 I also calculate AICc, which is the second order derivative of AIC and cor-
 124 rect small sample size. When $size/40 < \text{number of parameters}$ and $size > 5$,
 125 I calculate AICc as a supplementary criteria of model selection. Here is the
 126 equation of AICc (k is the number of free parameters and n is the sample
 127 size):

$$AIC_c = AIC + \frac{2k^2 + 2k}{n - k - 1} \quad (4)$$

128 **2.2.4 Plotting and analysis**

129 I draw line charts of AIC, BIC and R^2 which give a direct discription on 3
 130 models' fitting performance in bacterial growth data. Moreover, the smaller
 131 AIC and BIC are and the bigger R^2 is, the better model performs. Therefore
 132 I calculate the best performing model propotion in AIC, BIC and R^2 criteria
 133 and visualize them as pie charts. Although only part of datasets meet the
 134 condition of using AICc, I analyze it and summarize it with the other 3
 135 model comparison criteria in a table.

136 **3 Results**

137 Finally, 282 of 284 datasets are fitted successfully in all of cubic polynomial
 138 model, Logistic model and Gompertz model. Also, all of the model fitting
 139 lines are drawn in the point graph of every dataset. Here is an example

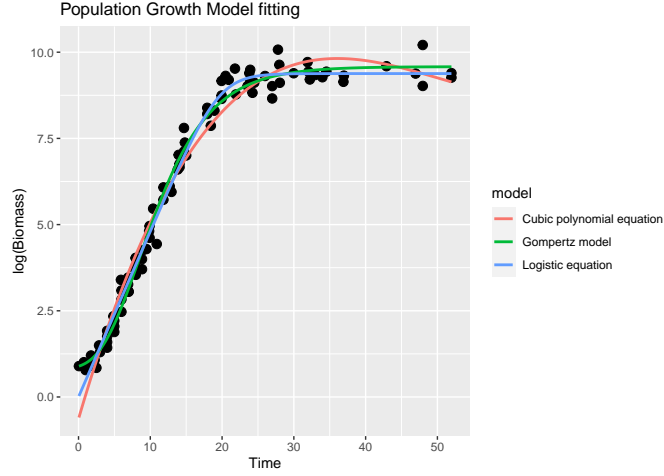


Figure 1: An example of 3 models fitting to a dataset

140 of the model fitting situation to the last dataset (Figure 1). This dataset
 141 contains 89 records, which is a relative big sample size among 284 datasets,
 142 and the population growth curve is also reasonable. It is obvious that every
 143 model shows a good fit to the growth of microbes population, but only Gom-
 144 pertz model catches the lag phase of microbe growth. Therefore, it seems
 145 like Gompertz model is the best model to fit this dataset. However, there
 146 are some more scientific and reasonable model selection criterias consider-
 147 ing fit and complexity and giving a specific number to show model fitting
 148 performance, such as AIC, AICc, BIC and R^2 . Table 1 demonstrates exact
 149 number of datasets that every model shows the best performance in AIC,
 150 AICc, BIC and R^2 .

Table 1: The best fitting model under 4 criteria

Model	AIC	AICc	BIC	R^2
Gompertz	172	127	169	192
Logistic	54	112	59	29
Cubic	56	25	54	61

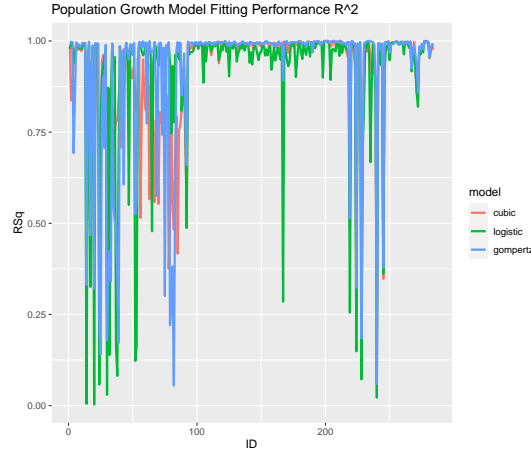


Figure 2: Model fitting performance R^2

151 3.1 R^2

152 R^2 is the residual sum of squares for a model, which is the simplest criterion
 153 to show the quality of fit. The more R^2 close to 1, the better model fits
 154 experiments data. As shown in Figure 2, cubic model shows general large
 155 R^2 , while both of Logistic model and Gompertz model appears to have some
 156 quite low R^2 . Calculated from Table 1, there are 68.1% datasets fitted best
 157 by Gompertz model, 21.6% datasets fitted best by cubic polynomial model
 158 and 10.3% datasets fitted best by Logistic model. However, maximizing R^2
 159 doesn't consider model complexity, neglecting the parsimony principle.



Figure 3: Model fitting performance AIC

3.2 AIC & AICc

The Akaike information criterion (AIC) is a model selection criterion that considers both fit and complexity among multiple models. When the sample size is small, AICc should be used to correct the bias. Gompertz model's AIC shows an overall relative low value compared with the other 2 models given Figure 3 line chart. In further best fitting model analysis, Gompertz model, cubic polynomial model and Logistic model fitting real microbes population growth data best accounts for 61%, 19.9% and 19.1% of 284 datasets respectively. However, when calculating AICc, there are 18 datasets' size less than or equal to 5, which can not be calculated AICc value. In 264 datasets, best fitting model proportion changes a lot, Gompertz model decreasing to 48.1%, cubic polynomial model decreasing to 9.5% and Logistic model increasing to 42.4%. But Gompertz model still accounts for the largest proportion of best fitting and shows general lower AICc value in Figure 4.

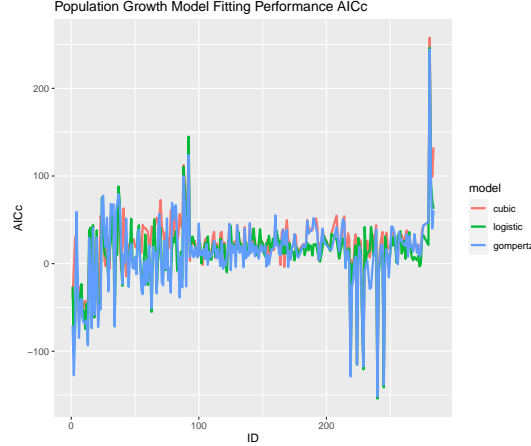


Figure 4: Model fitting performance AICc

3.3 BIC

BIC is a model selection criterion taking fit ,complexity and sample size into account. It prefers simpler models, especially sample size increasing. Three models' fitting performance in Bayesian information criterion (Figure 5) is similar with what AIC interprets, Gompertz model showing a globe lower BIC. The best performing model propotion in BIC is 59.9% for Gompertz model, 19.1% for cubic polynomial model and 20.9% for Logistic model.

4 Discussion

Population growth is a complex system affected by massive factors. What is a good method to deal with such a complicated system is setting up mathematical models. 3 criteria of a mathematical model is generality, realism and precision (LEVINS 1966). When establishing models, we should balance these 3 criteria according to our research objective. Strictly speaking, three models in this program are all empirical models instead of totally mechanistic models. The purpose of these models is to describe and quan-

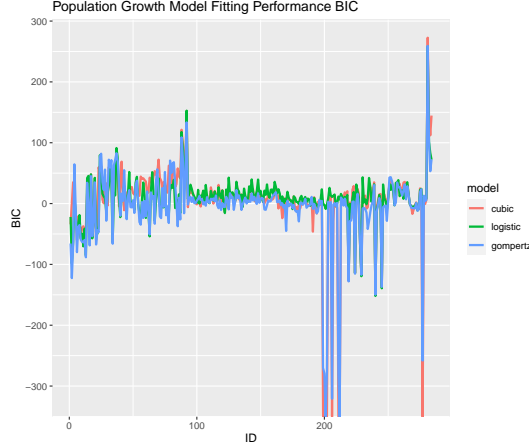


Figure 5: Model fitting performance BIC

189 tify experimentally observed patterns, but not explain why this particular
 190 pattern emerges (Peleg & Corradini 2011). Therefore, the mathematical
 191 convenience and good fit performance matter. Here I use AIC, BIC and R^2
 192 to show statistical measure of model fitting, and try to answer whether time
 193 lag is worthy of introduced in microbes population growth model.

194

195 In the data management, I deleted a dataset with too high propotion of
 196 negative biomass. Therefore, the other 284 datasets are fitted by cubic
 197 polynomial model, Logistic model and Gompertz model after deleted neg-
 198 ative biomass and time. Ultimately, 282 datasets are successfully fitted in
 199 all of these 3 models, and the fitting lines seem rational and well-fitted to
 200 describe microbes population growth. R^2 is used to show the goodness of
 201 fit, while AIC, AICc and BIC are used as model selection criteria, consid-
 202 ering both fit and model complexity. Cubic polynomial model seems to fit
 203 microbes experiments data well with general high R^2 value while Gompertz
 204 model shows the highest R^2 value in 192 datasets. For model selection,
 205 Gompertz model shows the lowest AIC and BIC value in majority of 282

206 datasets, which means although BIC favor simpler model and have correct
 207 sample size bias, Gompertz model still performs better than the other 2
 208 models. When considering AICc, it is the same case. In conclusion, Gom-
 209 pertz model is the best of these 3 models to describe microbes population
 210 growth.

211

212 The success of Gompertz model demonstrates at least 2 enlightenments.
 213 Firstly, mechanistic model is better than totally phenomenological linear
 214 model when the number of parameters are the same. Time lag is a worth-
 215 while parameter to be added into microbes population growth data which
 216 can improve model fitting performance significantly. The biological reason
 217 of this phenomenon may be the transition behavior of microbes at the end
 218 of lag phase of population growth process (Verhulst et al. 2011). Although
 219 the value of parameter time lag might not be equal to the biological meaning
 220 of lag phase duration.

221

222 Besides, there are many other population growth model not involved in
 223 this program. Such as Baranyi model (BARANYI et al. 1993), three-phase
 224 model (Buchanan et al. 1997), totally and truly mechanistic model on cells
 225 level (Peleg & Corradini 2011) and so on. Baranyi model is also a widely used
 226 model to describe population growth. However, instead of adding time lag
 227 parameter, it added the initial physiological state of cells h_0 compared with
 228 Gompertz model. Furthermore, this model used an equation $h_0 = t_{lag}r_{max}$,
 229 appearing in the solution of rate equation, to calculate time lag. The relative
 230 high correlation between r_{max} and h_0 results in difficulty of estimating the
 231 parameters value of Baranyi model (Grijpspeerdt & Vanrolleghem 1999). But
 232 a more district experiments design can guarantee the precise of parameter
 233 estimate. Also the benefits of introducing h_0 in model should be calculated
 234 in the future.

235

236 Whitting specified models into 3 levels (Whiting 1995). All models men-
 237 tioned above are all level one model that only describe changes of microbial

238 numbers versus time. Level 2 models demonstrates effect of environment on
 239 parameters in level 1. What level 3 models do is combining level 1 and level
 240 2 models and calculating microbial behavior under changed environment
 241 condition. We should choose suitable models according to our objective.
 242 Such as cubic polynomial model can catch death phase, that may be the
 243 reason why in some datasets cubic model shows the best fit. If we need to
 244 predict different temperature population growth (level 3 models), Gompertz
 245 model has some shortcomes and Baranyi model is the better choice of level
 246 1 model (Peleg & Corradini 2011, Silva et al. 2018). When using models,
 247 we ought to keep in mind that an appropriate model to the question is more
 248 important than model fitting performance. The validation of a model is that
 249 it generates good testable hypotheses relevent to crucial problems.

250 References

- 251 Aldila, D. & Seno, H. (2019), ‘A population dynamics model of mosquito-
 252 borne disease transmission, focusing on mosquitoes’ biased distribu-
 253 tion and mosquito repellent use’, *Bulletin of mathematical biology*
 254 **81**(12), 4977–5008.
- 255 BARANYI, J., ROBERTS, T. A. & MCCLURE, P. (1993), ‘A non-
 256 autonomous differential equation to model bacterial growth’, *Food mi-
 257 crobiology* **10**(1), 43–59.
- 258 Bernhardt, J. R., Sunday, J. M. & O’Connor, M. I. (2018), ‘Metabolic the-
 259 ory and the temperature-size rule explain the temperature dependence of
 260 population carrying capacity’, *The American naturalist* **192**(6), 687–697.
- 261 Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M.,
 262 Comita, L., Crone, E., Cubaynes, S., Davies, T., de Valpine, P., Ford,
 263 J., Gimenez, O., Kéry, M., Kim, E. J., Lennert-Cody, C., Magnusson,
 264 A., Martell, S., Nash, J., Nielsen, A., Regetz, J., Skaug, H. & Zipkin, E.
 265 (2013), ‘Strategies for fitting nonlinear ecological models in r, ad model

- 266 builder, and bugs', *Methods in Ecology and Evolution* **4**(6), 501–512.
 267 **URL:** [https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-](https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12044)
 268 [210X.12044](https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12044)
- 269 Buchanan, R., Whiting, R. & Damert, W. (1997), 'When is simple good
 270 enough: a comparison of the gompertz, baranyi, and three-phase linear
 271 models for fitting bacterial growth curves', *Food microbiology* **14**(4), 313–
 272 326.
- 273 Burnham, K. P. (2002), *Model selection and multi-model inference : a prac-*
 274 *tical information-theoretic approach*, 2nd ed. edn, Springer, New York ;
 275 London.
- 276 Gamito, S. (1998), 'Growth models and their use in ecological modelling:
 277 an application to a fish population', *Ecological Modelling* **113**(1), 83 – 94.
 278 **URL:** <http://www.sciencedirect.com/science/article/pii/S0304380098001367>
- 279 Grijspeerdt, K. & Vanrolleghem, P. (1999), 'Estimating the parameters of
 280 the baranyi model for bacterial growth', *Food microbiology* **16**(6), 593–
 281 605.
- 282 Hendry, A. P. & Green, D. M. (2017), 'Eco-evolutionary dynamics in cold
 283 blood', *Copeia* **105**(3), 441–450.
- 284 Johnson, J. B. & Omland, K. S. (2004), 'Model selection in ecology and
 285 evolution', *Trends in Ecology and Evolution* **19**(2), 101 – 108.
 286 **URL:** <http://www.sciencedirect.com/science/article/pii/S0169534703003458>
- 287 LEVINS, R. (1966), 'The strategy of model building in population biology',
 288 *American scientist* **54**(4), 421–431.
- 289 Peleg, M. & Corradini, M. G. (2011), 'Microbial growth curves: What the
 290 models tell us and what they cannot', *Critical Reviews in Food Science*
 291 *and Nutrition* **51**(10), 917–945. PMID: 21955092.
 292 **URL:** <https://doi.org/10.1080/10408398.2011.570463>

- 293 Peleg, M. & Shetty, D. K. (1997), ‘Modeling microbial populations with
294 the original and modified versions of the continuous and discrete logistic
295 equations’, *Critical Reviews in Food Science and Nutrition* **37**(5), 471–
296 490. PMID: 9315435.
297 **URL:** <https://doi.org/10.1080/10408399709527785>
- 298 Sarker, S. & Wiltshire, K. H. (2017), ‘Phytoplankton carrying capacity: Is
299 this a viable concept for coastal seas?’, *Ocean and Coastal Management*
300 **148**, 1 – 8.
301 **URL:** <http://www.sciencedirect.com/science/article/pii/S0964569117301977>
- 302 Silva, A. P. R. d., Longhi, D. A. & Dalcanton, F. (2018), ‘Modelling
303 the growth of lactic acid bacteria at different temperatures’, *Brazilian*
304 *Archives of Biology and Technology* **61**.
305 **URL:** <https://doi.org/10.1590/1678-4324-2018160159>
- 306 Verhulst, A., Cappuyns, A., Van Derlinden, E., Bernaerts, K. & Van Impe,
307 J. (2011), ‘Analysis of the lag phase to exponential growth transition
308 by incorporating inoculum characteristics’, *Food microbiology* **28**(4), 656–
309 666.
- 310 Whiting, R. C. (1995), ‘Microbial modeling in foods’, *Critical Reviews in*
311 *Food Science and Nutrition* **35**(6), 467–494.
312 **URL:** <https://doi.org/10.1080/10408399509527711>
- 313 Zwietering, M. H., Jongenburger, I., Rombouts, F. M. & van ’t Riet, K.
314 (1990), ‘Modeling of the bacterial growth curve’, *Applied and Environ-*
315 *mental Microbiology* **56**(6), 1875–1881.