

IMPERIAL COLLEGE LONDON

DEPARTMENT OF LIFE SCIENCE

The success of Gompertz models in fitting population growth data

Author:
Zongyi Hu

Supervisor:
Samraat Pawar

Date: 1/21/2021
Word Count: 2043

Abstract

Assessing the relationship between population size and time is of the essence in ecology and evolution field. In this project, I attempt to study this relationship by operating model fitting, the powerful and versatile and widely used technique. In this project, three widely used models: polynomial(cubic) model, logistic model and Gompertz model were chosen to be fitted using minpack.lm package R language which is based on the Levenberg-Marquardt algorithm. All three models are fitted successfully without any serious problem on several data sets obtained from published papers studying population growth of bacteria and phytoplankton under various experimental conditions in this project. The goodness-of-fit of each model were compared by making inference through comparing the $AICc$, AIC and BIC values. After comparison, the Gompertz model gave the overall best fit for the data, generally because of its ability to capture the lag phase of the population growth. The cubic and logistic model gave similar performance.

1 Introduction

The objective of this project is to fitted Three models and quantitatively comparing the criterion of fitting effect get the better model. As population density arranges along the time playing an essential role in ecology functions and evolutionary processes, assessing the population kinetic is worthy. Model fitting, as a powerful and versatile approach applied in a wild variety of research analysis, gives us a way of addressing this analyses. By fitting models, estimating the relationship between variables, we can make further inference to explore the mechanism behind it.

Base on theory and subject knowledge, the cubic model, logistic model and Gompertz model(Zwietering et al. 1990) are chosen to be fitted. For using the 4-parameter Gompertz model, all the denpendent variable is log-transformed(Tjørve & Tjørve 2017). The models split the curves of bacterial population growth into three phases: the lag phase at the beginning and the stationary phase at the end, those two phases have the specific growth rate of zero, and the exponential phase where the bacterial population increases nearly linearly over time, which corresponded to the parameter: $r_m a x$ in model equations. In both logistic and Gompertz models, the $r_m a x$ is calculated by using the slope of the linear model fitted on the whole data set

if the size of the data set is less than 5, or else, the data points chosen for fitting the linear model in the logistic model is the points in the middle 60% time, and for Gompertz is the points within 20% - 75% population size.

Though the logistic model, which represents density-dependent growth(Eberhardt et al. 2008), is good enough of representing the self density restriction, which in equation:

$$\frac{dN_t}{dt} = rN_t(1 - \frac{N_t}{K}) \quad (1)$$

is restricted by the part $1 - N_t/K$. In reality, however, there is one more phenomenon noticeable is that the population does not growth promptly after getting into the new environment. So, to capture the lag phase, the more complicated growth model: Gompertz model was introduced in this project which is asymmetrical compared with the logistic model. The Gompertz model has one more parameter t_{lag} to represent the last time point before the population exponentially growth and from the Gompertz equation

$$N_t = N_0 + (N_{max} - N_0)e^{-e^{\frac{r_{max}exp(1)}{(N_{max}-N_0)log(10)}+1} \frac{t_{lag}-t}{(N_{max}-N_0)log(10)}+1}} \quad (2)$$

we can see that, N_0 , the lower asymptote, compresses the curve by lifting the lower asymptote without altering the upper asymptote: N_{max} .(Tjørve & Tjørve 2017)

After fitting models, we can see relatively obscure difference of the fitting effect among three models through visually comparison, so the model selection criteria needed in this project. Rather than the conventional null hypothesis testing approach, the *AICc*, *AIC* and *BIC*(Johnson & Omland 2004) were adopted as model selection criteria. In which several competing hypotheses are simultaneously confronted and among model selection criteria, *AIC* and *AICc* are generally favoured because it has its foundation in Kullback–Leibler information theory (Anderson & Burnham 2004).

2 Methods

2.1 Computing Tools

The fitting scripts are written in R, because R is wildly used in academic and has an extensive library of tools for data and database manipulation and wrangling resource, like minpack package used in this report. And R

has many good quality of data visualization which makes it suitable for this project. Besides, R is open source and is not severely restricted to operating systems, which give another reason to choose R as analysis tool for this project. Python is used to build the whole workflow and generate the latex report.

2.2 Data Management

The data analysed in this project is collected from published works(Bae et al. 2014, Bernhardt et al. 2018, Galarz et al. 2016, Gill & DeLacy 1991, Phillips & Griffiths 1987, Roth & Wheaton 1962, Silva et al. 2018, Sivonen 1990, Stannard et al. 1985, Zwietering et al. 1994), study microbes and phytoplankton, contain populational growth data.

What concerns in this project are time and population size variables, so the ID, contains the species, experiment conditions and citations, was inserted in the data processing, and the data points have negative population size which does not have any biological meaning were deleted.

Then the population size in each data set was log-transformed for better analysing when the population growth is still in the lag phase with too small size and comparing the Gompertz model, which models the log-transformed dependent variables, with other two models when calculating the $AICc$, AIC and BIC value of each model in the identical standard.

So far, the 285 data sets were ready to process the model fitting.

2.3 Models and Equations

2.3.1 Models

The linear cubic model, logistic model and Gompertz model(Zwietering et al. 1990) are chosen in this project to evaluate the data set respectively. The models split the curves of bacterial population growth into three phases: the lag phase at the beginning and the stationary phase at the end, those two phases have the specific growth rate of zero, and the exponential phase where the bacterial population increases nearly linearly over time, which corresponded to the parameter: r_{max} in model equations. Though the logistic model, which represents density-dependent growth(Eberhardt et al. 2008), is good enough of representing the self density restriction, in reality, how-

ever, there is one more phenomenon noticeable is that the population does not growth promptly after getting into the new environment. So, to capture the lag phase, the more complicated growth model: Gompertz model was introduced in this project which is asymmetrical compared with the logistic model and has one more parameter t_{lag} to represent the last time point before the population exponentially growth. The other two parameters used in this model fitting is N_0 and N_{max} , respectively represents the logarithm of minimum and maximum population size in data.

2.3.2 Parameter Estimation

All of the parameters in cubic equation: Equation3

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

a,b,c,d do not have biological meaning, and for it is a linear model, we do not have to estimate the starting value. For logistic and Gompertz model, I get the preliminary starting value first and sampled those 1000 times with 0.2 factor to fit repeatedly in each data set.

2.4 Data Standardization (data.R)

I processed the data standardization by deleting the population size of negative numbers which do not have any biological meaning, and log-transformed the for better analysing when the population growth is still in the lag phase with too small size and comparing the Gompertz model, which models the log-transformed variables, with other two models when calculating the $AICc$, AIC and BIC value of each model in the identical standard.

2.5 models fitting on experimental data sets (Gompertz.R and Logistic.R)

I defined functions to process the model fitting, comparing the fitting effect, plotting and generating report. In which, several packages like ggplot2, and minpack are used. In processing the non-linear models it is possible to fail when fitting with too far-reach starting values, so within the function, the `tryCatch()` function is used to return the error and avoid stop the whole process. The visualization of the fitting is preseneted by plotting the actual

126 data points overlap with predicted lines to visualize the effects of the fitting,
127 which will be shown in the result section.

128 **2.6 Integrate the Whole Project(run_MiniProject.py)**

129 The whole process of the project is integrated into the single script and
130 generates the submission PDF file.

131 **2.7 Model Selection**

132 The comparison criteria in this project are $AICc$, AIC and BIC (Anderson & Burnham 2004). AIC and BIC were calculated by the inbuilt function
133 in R, $AICc$ was calculated by using the equation:
134

$$AICc = AIC + \frac{2K(K+1)}{n-K-1} \quad (4)$$

135 where n represents the number of data points, K is the number of parameters.
136 From the equation of calculating the $AICc$, we can see that if there
137 are more parameters in model equations, theoretically you will get higher
138 $AICc$ which means the worse performance of fitting effects. Even that, the
139 Gompertz model which has one more parameter is still consistently favoured
140 in this project.

141 **3 Results**

142 Generally, all the models were fitted without major problems and all the data
143 sets were fitted successfully by cubic and logistic models, 277 by Gompertz
144 model.

145 The typical fitted model visualizations look like (Figure1). From the
146 plot we can see all three models fitted well, the big difference between the
147 Gompertz model and the other 2 models are the ability of catch the lag
148 phrase. No big difference from visualization comparison, so that more precise
149 inference needs to be made by comparing the quantitative results, which in
150 this project are comparison criteria: $AICc$, AIC and BIC .

151 By comparing the $AICc$, AIC and BIC value of each model in all fitted
152 277 data sets, the Gompertz model tends to have lower criterion values with
153 winning frequencies: 144, 203 and 199, which shows that it is constantly

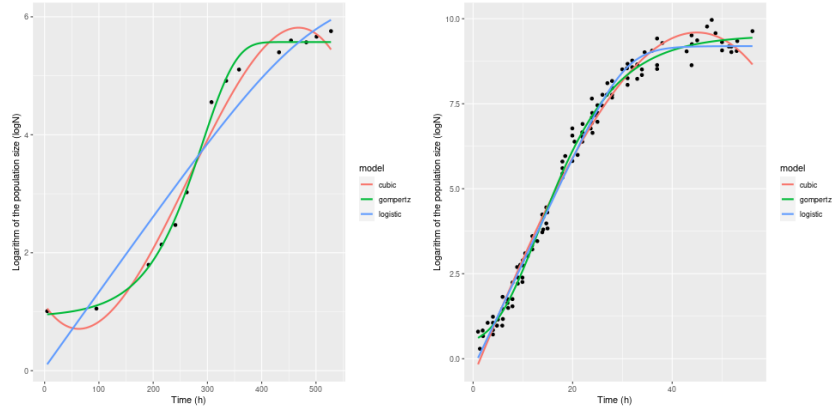


Figure 1: The general fitting effects visualization: two examples of fitting effects of cubic model, logistic model and Gompertz model on experimental data set. left graphic is the result of 68th data set, which experiment species is *Anthrobacter simplex* grown in TGE agar medium, right is 148th with *Lactobaciulus plantarum* grown in MRS medium.

154 better than cubic and logistic model with winning frequencies 19, 37, 36
 155 and 114, 37, 42 respectively. Also, from that we can see that the difference
 156 between cubic and logistic model are small. By all means, the Gompertz
 157 model gave the best fit to the data (Figure2).

158 4 Discussion

159 Is Gompertz model better than the cubic and logistic model? In this project,
 160 it is confident to say so. Theoretically, the model just has less parameters

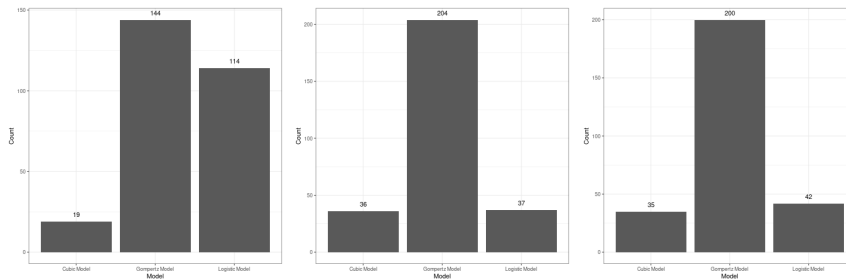


Figure 2: The winning frequency of the models: from left to right, the figures show the comparison on AIC_c , AIC and BIC values between 3 models.

161 should be considered first using, because it has more degrees of freedom and
 162 more stable since the parameters are less correlated, but the comparison of
 163 the AIC_c , AIC and BIC values indicated that even the Gompertz model
 164 has one more parameter than logistic model, it still outperformed. Through
 165 the result of AIC_c , because in this project, most data sets are small, so the
 166 calculation of the AIC_c besides AIC is necessary. It is calculated base on
 167 AIC with the equation4, which gives higher results with more parameters,
 168 even that is the case, the Gompertz model still performs better. The same
 169 situation happened when checking BIC , which favor simpler model and
 170 have correct sample size bias. Although the AIC is supposed to be used
 171 in large data sets, this project still calculated it as inference, the Gompertz
 172 model still wins. Gompertz model consistently preformed the best among
 173 the three models fitting on the data sets. So that, in this project, we can
 174 get the conclusion that the Gompertz model is the preferred model among
 175 the 3 cases.

176 The population usage of Gompertz model may because of its transfor-
 177 mation of the population size from normal to logarithm which makes its
 178 growth curve more linear in potential growth phase(Buchanan et al. 1997),
 179 also, the logarithm of the population size makes it more intuitive of biological
 180 meaning.

181 Although the linear cubic model in this project did the worst perfor-
 182 mance than the other two models, it can be used effectively in the absence
 183 of the stationary data(Buchanan et al. 1997). Its simplicity and flexibility
 184 give it the advantage to be used as a quick way of representing the good
 185 enough curve of the population growth without much effort.

186 The logistic model's relative failure, because of its disadvantage of catch-
 187 ing the lag phase compared with Gompertz model, while its relative vic-
 188 tory compared with cubic model, shows its ability of self density restric-
 189 tion(Eberhardt et al. 2008). Moreover the logistic model has one fewer
 190 parameter than do the cubic and Gompertz model and can be fitted to
 191 most of the data sets used in this project just without the ability to catch
 192 where growth had more than one phase but can be used to predict general
 193 growth(Balmer et al. 2012).

194 As known that the biotic and abiotic factors (Loreau 1998) both influ-
 195 ence the kinetic of the population size, which makes it a complex system

196 to be fitted with different circumstance. In this project, the data is all col-
 197 lected from bacteria and phytoplankton experiments we can fit and get the
 198 result that among three models: cubic, logistic and Gompertz, the Gompertz
 199 model stands out as the best model, while we can not infer this conclusion
 200 to other conditions. Generally, the mechanistic model performs better than
 201 only experimental model, but the biological meaning of the parameters in
 202 models is also needed to be concerned. There are already scholars trying
 203 to address this problem, Braillard, Pierre-Alain(Braillard 2010) mentioned
 204 that, design explanations should be considered as perfectly compatible with
 205 mechanistic explanation.

206 Further more, in equation2, it is obvious that fitting the Gompertz model
 207 is easier of meeting problems such as parameter identifiability problem??.
 208 Further analysis about the effective of this parameter needs to be operated
 209 to prove the worth of the effort of it. Then under different circumstances
 210 more appropriate and economic model can be chosen to use.

211 References

- 212 Anderson, D. & Burnham, K. (2004), ‘Model selection and multi-model
 213 inference’, *Second. NY: Springer-Verlag* **63**(2020), 10.
- 214 Bae, Y.-M., Zheng, L., Hyun, J.-E., Jung, K.-S., Heu, S. & Lee, S.-Y. (2014),
 215 ‘Growth characteristics and biofilm formation of various spoilage bacte-
 216 ria isolated from fresh produce’, *Journal of food science* **79**(10), M2072–
 217 M2080.
- 218 Balmer, N., Pleasence, P. & Nevill, A. (2012), ‘Evolution and revolution:
 219 Gauging the impact of technological and technical innovation on olympic
 220 performance’, *Journal of Sports Sciences* **30**(11), 1075–1083.
- 221 Bernhardt, J. R., Sunday, J. M. & O’Connor, M. I. (2018), ‘Metabolic the-
 222 ory and the temperature-size rule explain the temperature dependence of
 223 population carrying capacity’, *The American Naturalist* **192**(6), 687–697.
- 224 Braillard, P.-A. (2010), ‘Systems biology and the mechanistic framework’,
 225 *History and philosophy of the life sciences* pp. 43–62.

226 Buchanan, R., Whiting, R. & Damert, W. (1997), ‘When is simple good
227 enough: a comparison of the gompertz, baranyi, and three-phase linear
228 models for fitting bacterial growth curves’, *Food microbiology* **14**(4), 313–
229 326.

230 Eberhardt, L., Breiwick, J. & Demaster, D. (2008), ‘Analyzing population
231 growth curves’, *Oikos* **117**(8), 1240–1246.

232 Galarz, L. A., Fonseca, G. G. & Prentice, C. (2016), ‘Predicting bacterial
233 growth in raw, salted, and cooked chicken breast fillets during storage’,
234 *Food Science and Technology International* **22**(6), 461–474.

235 Gill, C. & DeLacy, K. (1991), ‘Growth of escherichia coli and salmonella
236 typhimurium on high-ph beef packed under vacuum or carbon dioxide’,
237 *International journal of food microbiology* **13**(1), 21–30.

238 Johnson, J. B. & Omland, K. S. (2004), ‘Model selection in ecology and
239 evolution’, *Trends in ecology & evolution* **19**(2), 101–108.

240 Loreau, M. (1998), ‘Biodiversity and ecosystem functioning: a mechanistic
241 model’, *Proceedings of the National Academy of Sciences* **95**(10), 5632–
242 5636.

243 Phillips, J. & Griffiths, M. (1987), ‘The relation between temperature and
244 growth of bacteria in dairy products’, *Food Microbiology* **4**(2), 173–185.

245 Roth, N. G. & Wheaton, R. B. (1962), ‘Continuity of psychrophilic and
246 mesophilic growth characteristics in the genus arthrobacter’, *Journal of*
247 *bacteriology* **83**(3), 551–555.

248 Silva, A. P. R. d., Longhi, D. A., Dalcanton, F. & Aragão, G. M. F. d. (2018),
249 ‘Modelling the growth of lactic acid bacteria at different temperatures’,
250 *Brazilian archives of biology and technology* **61**.

251 Sivonen, K. (1990), ‘Effects of light, temperature, nitrate, orthophosphate,
252 and bacteria on growth of and hepatotoxin production by oscillatoria
253 agardhii strains.’, *Applied and environmental microbiology* **56**(9), 2658–
254 2666.

- 255 Stannard, C., Williams, A. & Gibbs, P. (1985), ‘Temperature/growth re-
256 lationships for psychrotrophic food-spoilage bacteria’, *Food Microbiology*
257 **2**(2), 115–122.
- 258 Tjørve, K. M. & Tjørve, E. (2017), ‘The use of gompertz models in growth
259 analyses, and new gompertz-model approach: An addition to the unified-
260 richards family’, *PloS one* **12**(6), e0178691.
- 261 Zwietering, M., De Wit, J., Cuppers, H. & Van’t Riet, K. (1994), ‘Modeling
262 of bacterial growth with shifts in temperature’, *Applied and Environmen-
263 tal Microbiology* **60**(1), 204–213.
- 264 Zwietering, M., Jongenburger, I., Rombouts, F. & Van’t Riet, K. (1990),
265 ‘Modeling of the bacterial growth curve’, *Applied and environmental mi-
266 crobiology* **56**(6), 1875–1881.