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 filed. 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_max in model equations. In both logistic and Gompertz models, the r_max 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:

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

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

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

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, 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 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

$$N_t = a + bt + ct^2 + dt^3 \tag{3}$$

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

2.6 Integrate the Whole Project(`run_MiniProject.py`)

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

2.7 Model Selection

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

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

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

3 Results

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

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

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

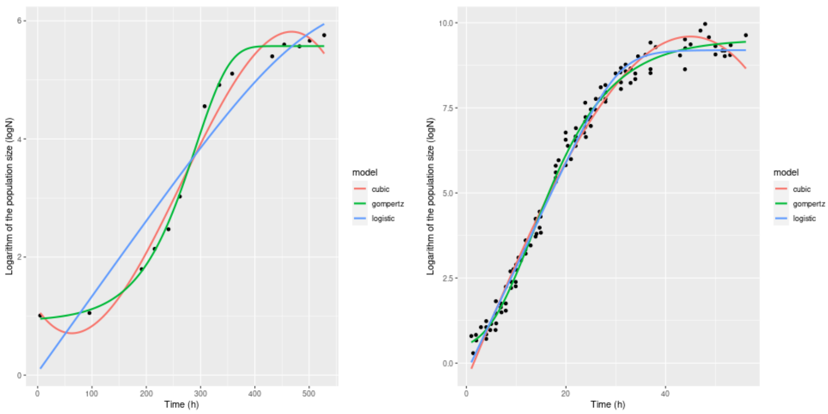

Figure 1: The general fitting effects visualization: two examples of fitting effects of cubic model, logistic model and Gompertz model on experimental data set. left graphic is the result of 68th data set, which experiment species is Anthrobacter simplex grown in TGE agar medium, right is 148th with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

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

4 Discussion

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

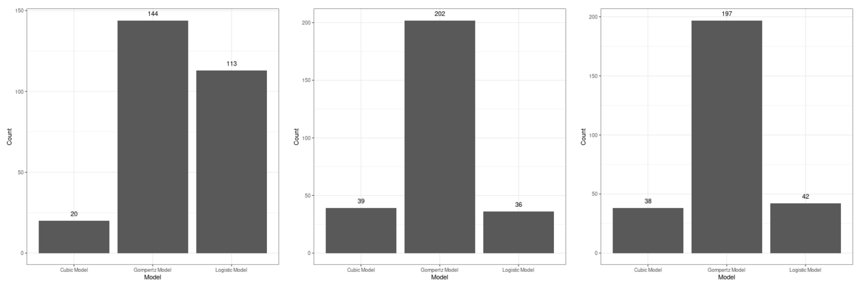

Figure 2: The winning frequency of the models: from left to right, the figures show the comparison on $AICc$, AIC and BIC values between 3 models.

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

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

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

The logistic model's relative failure, because of its disadvantage of catching the lag phase compared with Gompertz model, while its relative victory compared with cubic model, shows its ability of self density restriction(Eberhardt et al. 2008). Moreover the logistic model has one fewer parameter than do the cubic and Gompertz model and can be fitted to most of the data sets used in this project just without the ability to catch where growth had more than one phase but can be used to predict general growth(Balmer et al. 2012).

As known that the biotic and abiotic factors (Loreau 1998) both influence the kinetic of the population size, which makes it a complex system to be fitted with different circumstance. In this project, the data is all collected from bacteria and phytoplankton experiments we can fit and get the result that among three models: cubic, logistic and Gompertz, the Gompertz model stands out as the best model, while we can not infer this conclusion to other conditions. Generally, the mechanistic model performs better than only experimental model, but the biological meaning of the parameters in models is also needed to be concerned. There are already scholars trying to address this problem, Braillard, Pierre-Alain(Braillard 2010) mentioned that, design explanations should be considered as perfectly compatible with mechanistic explanation.

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

References

Anderson, D. & Burnham, K. (2004), 'Model selection and multi-model inference', *Second. NY: Springer-Verlag* **63**(2020), 10.

Bae, Y.-M., Zheng, L., Hyun, J.-E., Jung, K.-S., Heu, S. & Lee, S.-Y. (2014), 'Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce', *Journal of food science* **79**(10), M2072–M2080.

Balmer, N., Pleasence, P. & Nevill, A. (2012), 'Evolution and revolution: Gauging the impact of technological and technical innovation on olympic performance', *Journal of Sports Sciences* **30**(11), 1075–1083.

Bernhardt, J. R., Sunday, J. M. & O'Connor, M. I. (2018), 'Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity', *The American Naturalist* **192**(6), 687–697.

Braillard, P.-A. (2010), 'Systems biology and the mechanistic framework', *History and philosophy of the life sciences* pp. 43–62.

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.

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

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

Gill, C. & DeLacy, K. (1991), ‘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.

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

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

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

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

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

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

Stannard, C., Williams, A. & Gibbs, P. (1985), 'Temperature/growth relationships for psychrotrophic food-spoilage bacteria', *Food Microbiology* **2**(2), 115–122.

Tjørve, K. M. & Tjørve, E. (2017), 'The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family', *PloS one* **12**(6), e0178691.

Zwietering, M., De Wit, J., Cuppers, H. & Van't Riet, K. (1994), 'Modeling of bacterial growth with shifts in temperature', *Applied and Environmental Microbiology* **60**(1), 204–213.

Zwietering, M., Jongenburger, I., Rombouts, F. & Van't Riet, K. (1990), 'Modeling of the bacterial growth curve', *Applied and environmental microbiology* **56**(6), 1875–1881.