

Gompertz models out-perform logistic models for quantifying abundance data

Bowen Duan
College of Life Science

December 2, 2023

Abstract

In biological research, accurate quantification of biological abundance is critical to understanding population dynamics. My aim in this project was to compare the effectiveness of Gompertz and Logistic models in quantifying abundance data. I employed nonlinear least squares to fit both models and used the Akaike Informativeness Criterion (AIC) and the Bayesian Informativeness Criterion (BIC) as criteria for assessing the goodness of fit of the models. The data I got showed that the Gompertz model generally exhibited lower AIC and BIC values on several test datasets, suggesting that it outperforms the Logistic model in fitting abundance data.

1 Introduction

In biological and ecological studies, accurate quantification of the biological abundance of populations as a function of time is critical to understanding and predicting population dynamics. Preston (1948) indicates that relative species abundance is a good proxy for the diversity of species in an ecosystem. Therefore, finding effective models of population dynamics is particularly important for understanding population change.

Over the past decades, biologists have widely used a variety of mathematical models to describe and predict changes in population size. These models have provided important tools not only for theoretical ecology but also for practical applications in species conservation and natural resource management. Among these models, the Gompertz model proposed by Zwietering et al. (1990) and the Logistic model have attracted much attention because of their unique biological assumptions and mathematical properties. These models are widely used in biological and ecological studies for the prediction and analysis of population dynamics. However, Johnson and Omland (2004) shows that there is still controversy in the academic community about which model more accurately reflects the population dynamics in a particular ecological environment.

In recent years, model selection methods have been gradually introduced to research in ecology and evolutionary biology. Johnson and Omland (2004) shows that this approach involves the simultaneous data validation of multiple competing hypotheses to identify the best model or a set of models with more support. Model selection not only provides a way of evaluating and comparing multiple models, but also reflects how well different models fit the data. Research of Levins (1966) shows that there are three prevalent model simplification strategies the first is to sacrifice universality for realism and accuracy; the second is to sacrifice realism for universality and accuracy; and the third strategy is to sacrifice accuracy for realism and universality.

In this project I investigated the use of nonlinear least squares to fit Gompertz models and Logistic models, and used the Akaike Informativeness Criterion (AIC) and the Bayesian Informativeness Criterion (BIC) as the criteria for evaluating the goodness of fit of the models. The AIC and the BIC not only take into account not only how well the model fits the data,

but also the complexity of the model, thus providing a way to balance model simplicity and goodness-of-fit in model selection according to the paper of Johnson and Omland (2004).

2 Methods

2.1 General Workflow

First, the original dataset is divided in terms of IDs, then appropriate mathematical models are selected to fit to the data, then plotted against the model parameters and analysed appropriately to determine which model provides a better fit.

2.2 Data

Initial data are presented in tabular form and include information such as observation number (X), time, population biomass, temperature, time unit, biomass unit, bacterial species, medium, number of replicates, and cited literature. As I wanted to compare the models in terms of abundance. Therefore my focus for the data was on the relationship between abundance and time. In order to segment the dataset to make it more researchable and relevant, I categorised the dataset with IDs.

2.3 Models

I chose a total of 5 models. Three of them are linear models and two non-linear models. The non-linear models I chose are Logistic model and Gompertz model, and their mathematical expressions are as follows:

$$N_t = \frac{N_0 \cdot K \cdot e^{r \cdot t}}{K + N_0 \cdot (e^{r \cdot t} - 1)} \quad (1)$$

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

Logistic model and Gompertz model are two classic ecological models, which are widely used to describe the growth pattern of biological populations. logistic model assumes that the population stops growing after reaching a certain capacity, which is consistent with the real situation of limited resources. limited resources. This model allows the population to undergo an initial rapid growth phase when resources are sufficient, whereas the Gompertz model can be interpreted as having a time lag in the initial phase, which is more realistic because individuals within the population need time to grow during the initial growth phase.

2.4 Model Fitting

In this study, I applied a variety of statistical models to analyze the growth data of biological populations, including linear models, polynomial models, Logistic models, and Gompertz models. The application of these models aimed to understand the dynamics of population growth from various perspectives.

2.4.1 Linear and Polynomial Model Fitting

Considering the potential complexity of the data, I first explored the fitting effectiveness of linear and polynomial models for population growth data. The specific steps were as follows:

- Iteratively processed each unique identifier (ID) in the dataset, imposing restrictions on various parameters according to realistic conditions, and first filtered the data to examine the growth characteristics of each subset.
- Fitted linear, quadratic polynomial, and cubic polynomial models to each data subset to capture different trends and patterns in the data.
- Calculated the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and coefficient of determination (R-squared) for each fitted model to evaluate their respective fitting qualities.

2.4.2 Logistic Model Fitting

Furthermore, I employed the Logistic model to more accurately describe the saturation characteristics of population growth. This process included:

- Defined the standard Logistic growth function, taking into account the initial size of the population, the growth rate, and the environmental carrying capacity.
- Based on the slope of the linear model as the initial estimate of the growth rate, I logarithmically transformed the data to find the maximum slope, which was used as the maximum growth rate. The maximum and minimum values in the data for each ID were set as initial values for K and N_0 , respectively. Finally, the Non-Linear Least Squares (NLLS) method was used for fitting.
- Similarly, calculated the AIC, BIC, residual sum of squares (RSS), total sum of squares (TSS), and coefficient of determination for each Logistic model to assess the quality of fit.

2.4.3 Gompertz Model Fitting

Lastly, I used the Gompertz model to describe another pattern of population growth, especially effective in depicting the rapid growth phase in the early stages. The fitting process included:

- Defined the Gompertz model, considering factors such as the initial size of the population, maximum growth rate, lag time, and maximum population size. Here, the lag time was initially set using the maximum value of the second derivative.
- Also used Non-Linear Least Squares for fitting, with special handling of potential errors during the fitting process.
- Calculated the corresponding statistical measures, including AIC, BIC, RSS, TSS, and coefficient of determination, for each Gompertz model.

The fitting results of all models were meticulously recorded and exported in CSV format for subsequent analysis and comparison. These model fitting steps were implemented in the R programming language, particularly utilizing the ‘minpack.lm’ package for efficient non-linear fitting.

2.5 Model Comparison Methods

I chose AIC and BIC as the criteria for comparing the goodness of fit of the model inside the model fitted data. AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are two commonly used information criteria for making model comparisons. They are both based on the trade-off between the quality of fit of a statistical model on given data and the complexity of the model. The goal of both guidelines is to find a model that fits the data well but is not overly complex. The principle of AIC is to select the model with the smallest value

of AIC; the smaller the AIC, the better the model fits the data, but it also takes into account the complexity of the model. AIC is more effective for smaller sample sizes. Similar to AIC, BIC also maximizes the likelihood function plus a penalty for model complexity. the penalty term in BIC increases with sample size, so BIC prefers simpler models when the sample size is large. Therefore, the criterion for the choice of AIC and BIC is the smaller the better: the smaller the values of AIC and BIC, the better. The reason I did not choose Coefficient of Determination as a way to determine whether a model is good or bad is that Coefficient of Determination generally considers how linear the model is and does not determine whether the model has a causal relationship.

3 Results

In comparing the goodness of Logistic and Gompertz models we can use to compare the size of the corresponding AIC and BIC. Recall that in the methods section above it was mentioned that the smaller the better for both figures. According to my code will generate the AIC and BIC data from the statistics in the dataset, here is the corresponding example.

Table 1: Comparison of Logistic and Gompertz Models

	Logistic Model		Gompertz Model	
	AIC	BIC	AIC	BIC
Sample 1	-104.12	-98.79	-216.46	-209.80
Sample 2	-56.64	-51.31	-87.10	-80.44
Sample 3	-44.41	-39.08	-55.51	-48.85
Sample 4	-31.88	-26.28	-37.91	-30.91
Sample 5	-254.75	-249.72	-257.83	-251.54

This table shows the corresponding AIC and BIC values for the first five IDs, which clearly shows that the Gompertz model is smaller for both data, and therefore shows that Gompertz is a better fit for population abundance and time.

4 Discussion

Recall that the initial and most fundamental question was to explore which mathematical model would best fit the empirical dataset, therefore I finally decided to choose two nonlinear models i.e. Logistic model and Gompertz model to fit the dataset for exploration. Based on what I learned in the introduction section of the paper I learned that I can use AIC and BIC as metrics to discuss which model is a better fit. Therefore my idea of the process for the code section is to first filter the data in the original dataset, i.e. add constraints to the parameters to filter out the data that does not fit the actual biological significance. Subsequently, R was used to script the model fitting and the mathematical model was fitted using LM and NLLS. Finally plotting scripts were written to generate the fitted images.

I failed in plotting Gompertz, probably because I didn't find the right initial values. But for the Logistic model, we can first add one to the popbio and take the logarithm, then use the LM and find the maximum slope of growth as the initial r , then use the minimum and maximum data in the corresponding sub-datasets to represent the initial population abundance and the maximum environmental holding capacity. However, for the Gompertz model, the statistics AIC, BIC, and the coefficients of determination are not meaningful, so there is no way to plot them one by one using ID as the division.

In conclusion, this report examines the AIC and BIC values of the logistic and Gompertz models to investigate the model fit and concludes that Gompertz is better than the logistic model in fitting the population abundance. The actual reason for this is that it takes time for individual population to grow in the early stage of growth, so Gompertz with time lag fits better. This paper can be used as a reference for the idea of model fitting, but the improvement is that the value of AICc can be calculated to further determine the degree of model fitting, because according to research of Johnson and Omland (2004). AICc is more suitable for data sets with small data volume to conduct research.

Word Count

1873 words

References

- Johnson, J. B. and Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2):101–108.
- Levins, R. (1966). The strategy of model building in population biology. *American scientist*, 54(4):421–431.
- Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology*, 29(3):254–283.
- Zwietering, M. H., Jongenburger, I., Rombouts, F. M., and Van't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–1881.