

Comparative Assessment of Mechanistic and Phenomenological Models for Bacterial Growth

Yibin Li

Imperial College London
yl2524@ic.ac.uk

14/3/2025
Word Count: 2980

Abstract

Understanding bacterial population growth is critical in fields ranging from clinical microbiology to industrial biotechnology. This study compares mechanistic (logistic, modified Gompertz, Baranyi) and phenomenological (quadratic, cubic) models using 275 datasets spanning 45 bacterial species and multiple media. Model performance was assessed via AICc, BIC, and R^2 . The logistic model reliably fit curves lacking a pronounced lag phase, while the modified Gompertz model excelled for complete S-shaped patterns, accurately capturing lag, exponential, and stationary phases. Although the Baranyi model theoretically accommodates lag phases, it was susceptible to missing stages and parameter initialization, sometimes resulting in convergence failures. Polynomial models, though flexible in partially observed or declining growth phases, lacked biological interpretability. These findings underscore that no single model is universally optimal; mechanistic models work best when datasets include all growth stages, whereas phenomenological models may be more suitable for incomplete or atypical data. Future work should refine mechanistic frameworks to incorporate death-phase parameters and explore machine learning approaches to bolster predictive robustness.

1 Introduction

Population growth studies are fundamental in ecology, microbiology, and evolutionary biology. Even small changes in abundance can trigger significant ecosystem effects, such as prey surges driving trophic cascades or pathogenic increases elevating disease prevalence (de Bello et al., 2021). A minor shift in a single population can disproportionately impact ecosystem stability and function (Callaghan et al., 2024). Since population dynamics shape community structure, regulate ecosystem processes, and influence nutrient cycling and disease transmission, understanding growth patterns is essential.

A particularly well-documented example is bacterial populations, which, under optimal conditions, often follow Malthusian growth encompassing four distinct stages (Allen and Waclaw, 2018). They enter a lag phase before transitioning to an exponential phase of constant population expansion, demonstrating Malthusian principles. Eventually, resource depletion or the accumulation of by-products leads to a stationary phase with near-zero net growth. If adverse conditions persist, bacteria enter a death phase marked by declining viable cells (Gonzalez and Aranda, 2023).

Quantifying these stages of bacterial growth is not just an academic exercise. It has practical implications for fields ranging from clinical microbiology to industrial biotechnology. In clinical settings, understanding the proliferation rate of pathogenic bacteria can provide important information for developing therapeutic strategies (Váradi et al., 2017). Meanwhile, in industrial microbiology, optimizing bacterial growth can significantly increase the production of bioderived products such as enzymes, antibiotics, and other metabolites (Vithalani et al., 2024).

Mathematical models have been developed to predict bacterial growth by capturing key growth stages, estimating biologically meaningful parameters (e.g., intrinsic growth rate, carrying capacity), and enabling standardized comparisons across species or strains (López et al., 2004). By linking observed data to theoretical frameworks, these models enhance outcome prediction and inform microbiological experimental design. Despite their shared objective, they differ in theory and parameterization. Mechanistic models (e.g., Logistic or Modified Gompertz) incorporate resource constraints and rates of change, yielding biologically interpretable parameters (Wang and Guo, 2024). In contrast, phenomenological models, such as polynomial fitting, prioritize data description without integrating biological processes (Aguirre-Zapata et al., 2024). These distinctions affect model generalizability and the extent to which mechanistic insights can be derived.

In light of these considerations, this study raises the following research questions:

1. **“Given that mechanistic and phenomenological models have different theoretical foundations and parameter settings, which type of model more accurately captures and fits the dynamic changes in growth data overall?”**
2. **“Do different models assign varying degrees of emphasis or ‘weight’ to different stages of the growth curve?”**

2 Method

2.1 Data

The dataset `LogisticGrowthData.csv` consists of 4,387 rows of data containing 45 unique bacterial species, 18 different media, and 10 different data sources. Grouped by species, medium, temperature, and data source, the dataset contains 285 unique sets of bacterial growth data from 10 different papers. The raw data was filtered and screened, and 114 lines with no biological significance were removed. The final 275 sets of different data were used for subsequent analysis and mapping.

2.2 Model

In this study, I used five different models to fit the data. These include two Phenomenological models, Quadratic and Cubic Models, and three Mechanistic models, Logistic model, Modified Gompertz Model, and Baranyi model.

1. Quadratic Model

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

2. Cubic Model

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

Where N_t is the population at time t , t is the time. a , b , c and d are parameters that do not represent any growth mechanism.

The Logistic Growth Model characterizes population growth limited by resource availability (Tsoularis and Wallace, 2002). It is expressed as:

3. Logistic Growth Model

$$N_t = \frac{N_0 K e^{r_{\max} t}}{K + N_0 (e^{r_{\max} t} - 1)} \quad (3)$$

- N_0 : Initial population abundance
- K : Environmental carrying capacity (maximum population abundance)
- r_{\max} : maximum rate of growth

The Modified Gompertz Model flexibly describes microbial growth, capturing its slow start, rapid acceleration, and gradual plateau (Gil et al., 2006). It is given by:

4. Modified Gompertz Model

$$N_t = N_0 + (K - N_0) \exp\left(-\exp\left(\frac{r_{\max} e (t_{\text{lag}} - t)}{(K - N_0) \ln(10)} + 1\right)\right) \quad (4)$$

- N_0 : Initial population abundance
- K : Environmental carrying capacity (maximum population abundance)
- r_{\max} : maximum rate of growth
- t_{lag} : Lag time before exponential growth stage

The Baranyi Model provides a seamless transition from initial adaptation to exponential growth (Grijspeerdt and Vanrolleghem, 1999). It is expressed as:

5. Baranyi Growth Model

$$A_t = t + \frac{1}{r_{\max}} \ln\left(\exp(-r_{\max} t) + \exp(-h_0) - \exp(-r_{\max} t - h_0)\right), \quad (5)$$

$$N_t = N_0 + r_{\max} A(t) - \ln\left(1 + \frac{\exp(r_{\max} A(t)) - 1}{\exp(K - N_0)}\right). \quad (6)$$

- N_0 : Initial population abundance
- K : Environmental carrying capacity (maximum population abundance)
- r_{\max} : maximum rate of growth
- h_0 : Physiological adaptation parameter

Baranyi suggested that $v = r_{\max}$, $m = 1$, and this method will also be adopted in our study. Parameter h_0 can be derived from r_{\max} (Baranyi and Roberts, 1994).

Here are the Statistical Metrics

1. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_i (y_{\text{obs},i} - y_{\text{fit},i})^2}{\sum_i (y_{\text{obs},i} - \bar{y}_{\text{obs}})^2} \quad (7)$$

- $y_{\text{obs},i}$: The i first observation
- $y_{\text{fit},i}$: The i model prediction value

- \bar{y}_{obs} : The average of the observed values

2. Corrected Akaike Information Criterion (AICc)

$$AIC = n \ln\left(\frac{RSS}{n}\right) + 2k, \quad (8)$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}, \quad (9)$$

$$RSS = \sum_i (y_{\text{obs},i} - y_{\text{fit},i})^2. \quad (10)$$

- n : Number of data points
- k : Number of model parameters

3. Bayesian Information Criterion (BIC)

$$BIC = n \ln\left(\frac{RSS}{n}\right) + k \ln(n). \quad (11)$$

- n : Number of data points
- k : Number of model parameters

4. Akaike Weights

$$w_i = \frac{\exp(-0.5 \Delta_i)}{\sum_j \exp(-0.5 \Delta_j)}, \quad \Delta_i = AICc_i - AICc_{\min}. \quad (12)$$

- $AICc_{\min}$: The smallest AICc of all candidate models
- w_i : The relative likelihood that a given model i becomes the "best model"

2.3 Model Fitting

In this experiment, the model's parameter definition and the dataset selection were aligned with practical significance. To avoid cases where the number of samples was smaller than the number of parameters, we selected datasets with more than five samples for analysis. The nonlinear model fitted the data using nonlinear least squares (NLLS). To improve the stability and accuracy of nonlinear fitting, we conducted 50 iterative analyses on each dataset, evaluated the model under

different starting value states, and selected the starting value corresponding to the highest R^2 as the best starting value of the dataset. The selection range of the initial value was defined as the initial value as the estimated value, and the value was taken by the regular distribution disturbance to avoid the local optimal solution.

In terms of model evaluation, due to the small number of samples in the dataset, AICc (9) was selected to evaluate the model's complexity and goodness of fit. BIC (11) selected a model that best fits the existing data from a fitting perspective. Akaike weight (12) measured the relative fitness between different models. The coefficient of determination R^2 (7) was used to measure the goodness of fit of an independent model.

2.4 Computing Tools

Data processing is carried out in R (version 4.4.2). The package `readr` was used to read and write table data. The package `dplyr` was used for data manipulation and transformation. Data fitting and plotting were performed in Python (version 3.12). `Pandas` processed tabular data (such as CSV files). `NumPy` provided efficient numerical computation and array manipulation. `Lmfit` was used to fit the linear model and the nonlinear model with NLLS. `Matplotlib.pyplot` was used to generate high-quality graphics. `SciPy.stats` provided several probability distributions and statistical functions. `RollingOLS` was a rolling window ordinary least squares (OLS) regression tool provided by `statsmodels` for calculating local linear regression on time series data. `Statsmodels` provided statistical models and econometrics tools to connect the above scripts into a clean, repeatable working environment. Running R and Python scripts through bash, and compile LaTeX scripts with references.

3 Results

Table 1: Model Comparison

Model	Convergence Rate (%)	Best selected AICc	Best selected BIC	Proportion of $0.7 < R^2 < 1$ (%)
Quadratic	100	54	21	77.8
Cubic	100	27	44	91.3
Modified Gompertz	100	185	186	94.5
Logistics	55.6	46	16	17
Baranyi	27.2	2	7	96

The Quadratic, Cubic, and Modified Gompertz model converged at 100% and fitted 275 bacterial growth curves, while Baranyi had the lowest convergence rate, 27.2%. Apart from the Quadratic model, the frequency of $0.7 < R^2 < 1$ in the remaining models accounted for more than 90% of the

successfully fitted datasets. The highest percentage of R^2 was achieved by the Baranyi model at 96%.

The AICc and BIC results indicated that the Modified Gompertz model adequately fitted most datasets. According to the AICc results, the Quadratic model was identified as the second optimal model in all fitted datasets 54 times, followed by the Logistics model 46 times.

In the BIC results, the Cubic model was rated as the second model in all fitted datasets 46 times, followed by the Quadratic model 21 times. The Baranyi model was rated by AICc and BIC as the best model with the lowest number of times, 2 times and 7 times respectively.

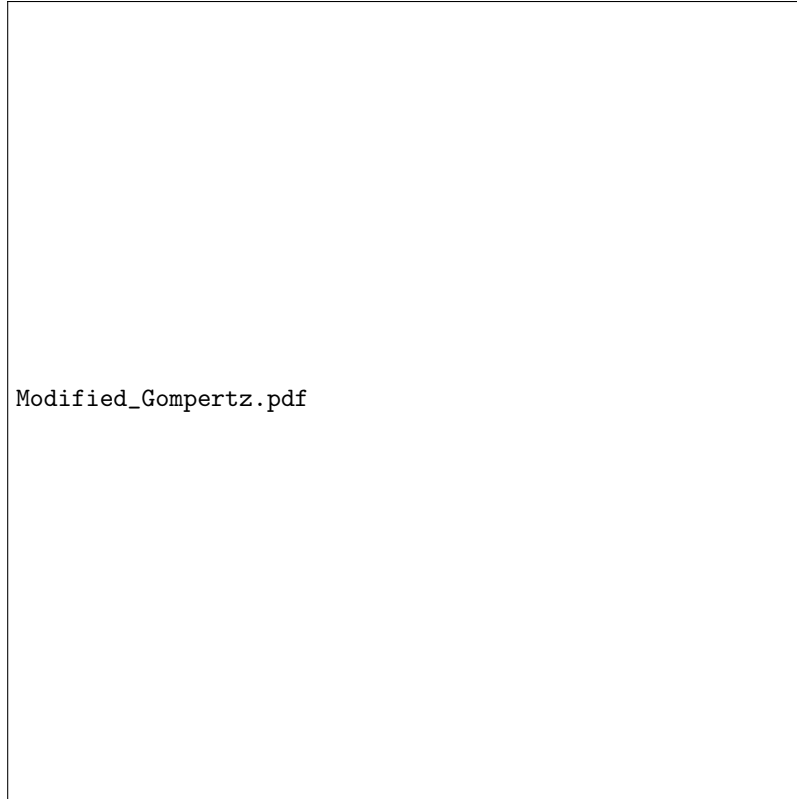


Figure 1: The best subset for Modified Gompertz model. This figure shows that Modified Gompertz model focuses more on the complete S-shaped growth curve with lag, exponential and stationary phase.



Figure 2: The best subset for Logistic model. This figure shows that logistic model focuses more on the growth curve with no significant obvious lag phase.

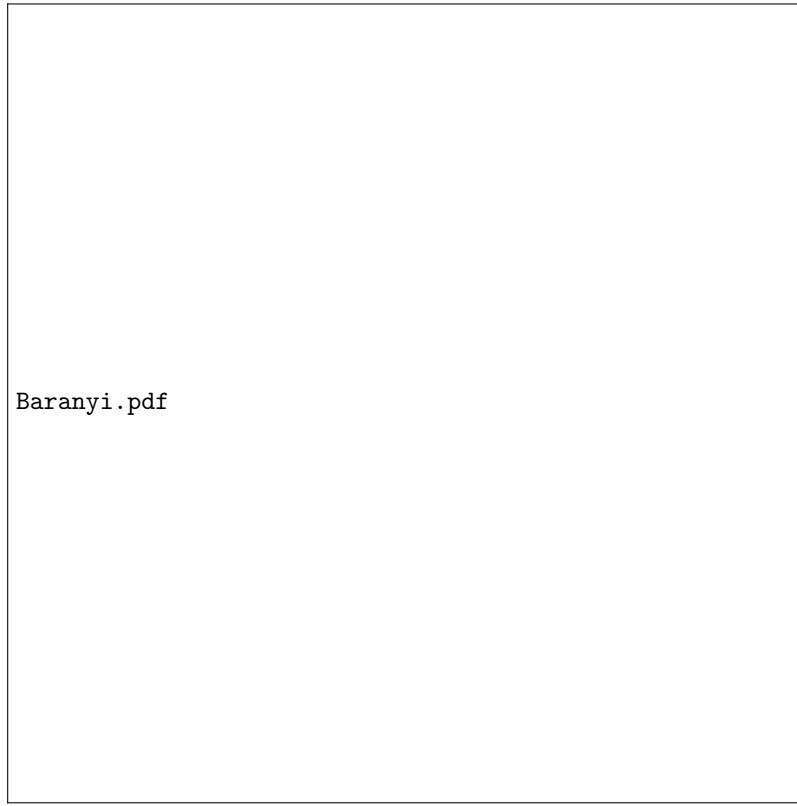


Figure 3: The best subset for Baranyi model. This figure shows that Quadratic model focuses more on the growth curve with early exponential phase.

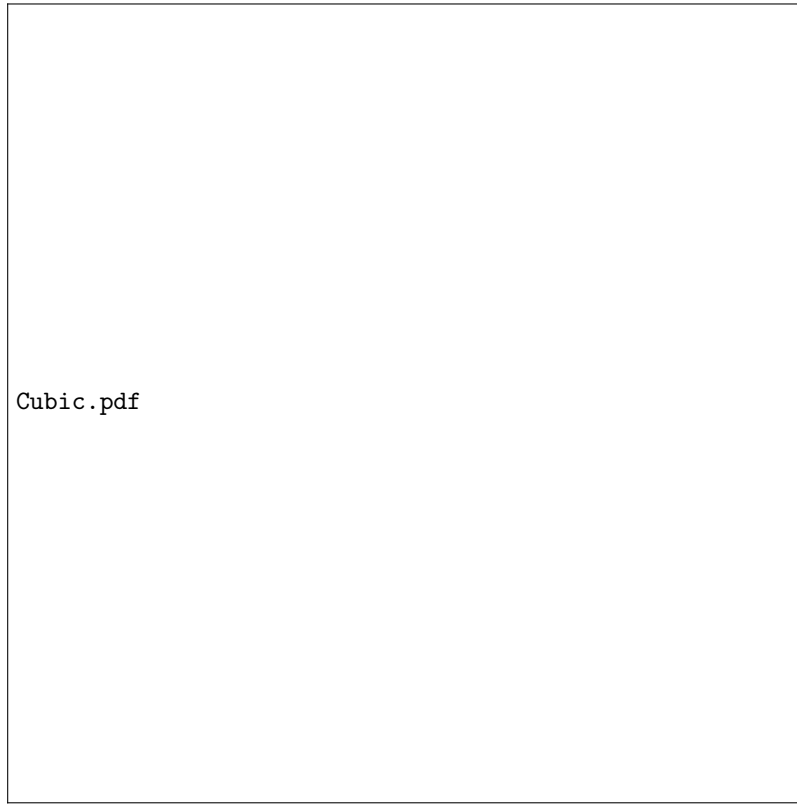


Figure 4: The best subset for Cubic model. This figure shows that Cubic model focuses more on the growth curve with death phase.

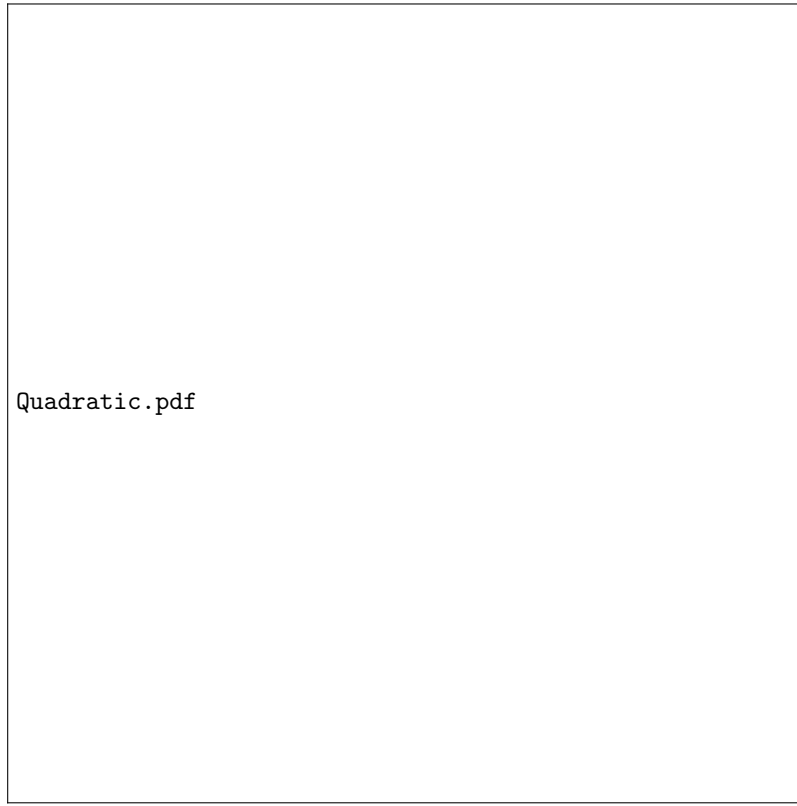


Figure 5: The best subset for Quadratic model. This figure shows that Baranyi model focuses more on the growth curve with lag phase.

The dataset that best fits each model was selected through Akaike Weight. The Figure contained the growth curve generated after the model fitted the data. The growth curves of different models were placed on the same figure for easy observation of fitting results.

4 Discussion

4.1 More appropriate model

Mechanistic models generally outperform phenomenological models in simulating bacterial growth by incorporating biological principles, such as resource constraints and lag phases, resulting in more accurate S-shaped curves. Their parameters provide ecological insights, enabling extrapolation beyond observed data and improving estimates of growth rates and carrying capacity. In contrast, phenomenological models, particularly Quadratic and Cubic polynomials, lack biological structure, leading to poorer extrapolation and higher AICc/BIC values.

However, mechanistic models like the Baranyi model can struggle or fail to converge when key growth phases (e.g., exponential or stationary) are missing or when sample sizes are small. Incomplete growth curves hinder reliable estimation of parameters such as carrying capacity and lag time, while sparse data complicates nonlinear fitting, producing unstable or unrealistic estimates. Poor fits (e.g., low R^2 or unstable curves) often stem from these limitations. Thus, while mechanistic models are preferred for datasets covering most growth stages, their advantages diminish with partial or limited data.

4.2 Varying degrees of emphasis

Figure 1 illustrates the characteristic S-shaped growth curve of the Modified Gompertz model, encompassing lag, exponential, and stationary phases. The Modified Gompertz model accurately captures both the initial and stationary lag phases. Compared to the polynomial curve, which deviates during the exponential phase, the Modified Gompertz model provides a superior fit by aligning lag times with the initial slow startup period. The Modified Gompertz model further enhances accuracy by incorporating biologically relevant parameters.

Figure 2 depicts the optimal Logistic model growth curve, where the population transitions directly into exponential growth without a clear lag phase. Under these conditions, the Modified Gompertz model, which includes lag-based parameters, proves unsuitable. This suggests that incorporating a lag parameter may be unnecessary or even counterproductive when no adaptation period is observed. The Logistic model's simpler structure suffices for datasets lacking a lag phase.

Figure 3 presents the best-fitting Baranyi model curve for a dataset with a distinct lag phase preceding exponential growth. The model's adjustment function $A(t)$ effectively simulates cellular adaptation, ensuring the curve remains low during the lag period and aligns with observed exponential growth. In contrast, the Logistic and Modified Gompertz models initiate growth prematurely, while the phenomenological model lacks the flexibility to accommodate a gradual onset. By explicitly accounting for microbial adaptation through parameters such as h_0 (initial physiological state), the Baranyi model emphasizes the lag phase more effectively than the Logistic or Modified Gompertz models.

Figure 4 presents the best-fitting Cubic model growth curve, capturing an initial population in-

crease, a plateau, and a subsequent decline. Unlike the Modified Gompertz model, which assumes a stable carrying capacity and cannot represent the mortality phase, the Cubic model’s mathematical flexibility allows it to reflect population decline accurately. The figure demonstrates that while the Cubic model tracks the post-peak decrease, the mechanistic model remains near the stationary phase, failing to capture mortality. This highlights the limitations of standard S-shaped models, which do not account for population decline. Despite being phenomenological, the Cubic model’s adaptability makes it preferable when data include a clear death phase.

Figure 5 shows the best-fitting Quadratic model growth curve. When data primarily represent the exponential growth stage, the Quadratic model outperforms mechanistic models, which may prematurely level off. Unlike the Modified Gompertz model, which underestimates later data, polynomial models, including the Quadratic and Cubic approaches, continue increasing through the observed points. These results suggest that polynomial models provide a more effective local approximation when data capture only a portion of the growth trajectory, particularly during the exponential phase.

Different modeling approaches emphasize distinct growth phases. The Baranyi model is optimal for datasets with a pronounced lag phase, while the Logistic and Modified Gompertz models effectively capture overall growth dynamics, excluding mortality. The Cubic model best represents population decline, whereas Quadratic models are useful for datasets covering only specific growth stages, preventing overfitting when mechanistic models estimate unobserved parameters. When data primarily include the exponential phase or lack a stationary phase, Quadratic fits often outperform S-shaped models by flexibly approximating growth trends without requiring a complete trajectory. Polynomial models further provide robustness by accommodating turning points across diverse data patterns.

However, polynomial models have limited explanatory power. While they allow interpolation within observed data, they are unreliable for predicting carrying capacity or lag time, and their coefficients lack biological interpretability. Thus, polynomial models serve as useful descriptive tools for incomplete datasets but are unsuitable for capturing full growth dynamics.

4.3 R^2

R^2 was not used for cross-model comparisons in this study due to its limitations in nonlinear and varying-complexity models. Since R^2 does not penalize additional parameters, more complex models may achieve similar or higher values, even when extra parameters lack practical significance. Some models reached near-1.0 R^2 yet failed to capture later growth stages, demonstrating that a single statistic can obscure key discrepancies. Additionally, error distribution across phases can distort R^2 , as a model may fit the exponential phase well while poorly representing the stationary phase. Thus, R^2 alone is unreliable for selecting a growth-curve model and was used here solely to assess model fit within individual datasets.

4.4 AICc or BIC

The ranking of models—determined by how often each is deemed best—varies depending on whether AICc or BIC is used, reflecting their distinct evaluation criteria. AICc prioritizes goodness of fit while imposing relatively modest penalties for additional parameters, whereas BIC applies stricter penalties, favoring simpler models in limited datasets. Despite lacking biological interpretability, the Cubic model performs well under BIC due to its mathematical flexibility, while complex S-shaped models are heavily penalized. The Baranyi model, highly sensitive to initial parameter values, exhibits instability and ranks lowest under both criteria. These results suggest the dataset captures fairly complete growth stages, including occasional decline phases.

4.5 Akaike weights

Akaike weights were used to estimate the relative probability of each model being the best in the candidate set, but they inherit certain limitations from AICc (Kolassa, 2011). While they provide a comparative measure of model quality, they do not fully account for complexities beyond AICc’s penalties. When models yield similar AICc values, their Akaike weights will also be similar, regardless of differences in complexity or interpretability. Relying solely on Akaike weights may lead to overfitting if minor improvements in fit favor unnecessarily complex models. Although Akaike weights adjust rankings by balancing fit and complexity, their interpretation requires caution. Future studies should integrate biological rationale and residual analysis to ensure model selection remains both statistically sound and biologically meaningful.

4.6 Initial parameter values

A major challenge in fitting nonlinear growth models is selecting appropriate initial parameter values (Paine et al., 2012). In this study, the success of the Logistic, Modified Gompertz, and Baranyi models depended on starting values being sufficiently close to the true parameters. Poor initial values led the Levenberg-Marquardt algorithm to either converge on a local minimum or fail entirely. To mitigate this, each dataset was tested with 50 random initial values within a permissible range, reducing erroneous convergence and improving the likelihood of a successful fit. However, optimal starting values were not guaranteed, particularly with limited or incomplete data. Future research would benefit from more comprehensive datasets and algorithms with automated or staged parameter adjustment to enhance robustness.

4.7 limitations

This study also highlights limitations in data quality and model assumptions. None of the models explicitly captured the death phase, suggesting that future approaches should incorporate dynamic switching or death-stage parameters for full-cycle predictions. Machine learning could further improve bacterial growth-curve modeling. Additionally, near-zero abundances resulted in extreme

logarithmic values, leading to numerical instabilities, particularly in the Baranyi model. These issues reveal sensitivity to noise and initial conditions, emphasizing the need for more robust algorithms, improved constraint optimization, or simplified model structures.

5 Conclusion

The choice of bacterial growth models significantly influences interpretation and predictive accuracy. Mechanistic models, such as the Modified Gompertz and Baranyi models, effectively capture S-shaped growth when all phases are present but perform poorly with missing stages due to their reliance on biologically meaningful parameters. In contrast, phenomenological models like Cubic and Quadratic polynomials offer flexible curve fitting but lack biological interpretability. No single model is universally optimal; selection should consider dataset structure, biological relevance, and computational efficiency. Parameter initialization is critical in nonlinear modeling, as poor starting values can cause convergence failures. Future research should refine mechanistic models to incorporate death-phase dynamics and explore hybrid approaches, including machine learning, to enhance robustness and predictive accuracy across clinical, industrial, and ecological applications.

References

- Aguirre-Zapata, E., Lema-Perez, L., Gomez-Echavarria, L., Botero-Castro, H., Maya, J. C., Chejne, F. and Alvarez, H. (2024), ‘Useful properties of phenomenological-based models’, *Ecological Modelling* **496**, 110850.
- Allen, R. J. and Waclaw, B. (2018), ‘Bacterial growth: a statistical physicist’s guide’, *Reports on Progress in Physics* **82**(1), 016601.
- Baranyi, J. and Roberts, T. A. (1994), ‘A dynamic approach to predicting bacterial growth in food’, *International Journal of Food Microbiology* **23**(3), 277–294.
- Callaghan, C. T., Santini, L., Spake, R. and Bowler, D. E. (2024), ‘Population abundance estimates in conservation and biodiversity research’, *Trends in Ecology & Evolution* **39**(6), 515–523.
- de Bello, F., Lavorel, S., Hallett, L. M., Valencia, E., Garnier, E., Roscher, C., Conti, L., Galland, T., Goberna, M., Májeková, M., Montesinos-Navarro, A., Pausas, J. G., Verdú, M., E-Vojtkó, A., Götzenberger, L. and Lepš, J. (2021), ‘Functional trait effects on ecosystem stability: assembling the jigsaw puzzle’, *Trends in Ecology & Evolution* **36**(9), 822–836.
- Gil, M. M., Brandão, T. R. and Silva, C. L. (2006), ‘A modified gompertz model to predict microbial inactivation under time-varying temperature conditions’, *Journal of Food Engineering* **76**(1), 89–94. Bugdeath.
URL: <https://www.sciencedirect.com/science/article/pii/S0260877405003389>
- Gonzalez, J. M. and Aranda, B. (2023), ‘Microbial growth under limiting conditions-future perspectives’, *Microorganisms* **11**(7).
- Grijpspeerdt, K. and Vanrolleghem, P. (1999), ‘Estimating the parameters of the baranyi model for bacterial growth’, *Food Microbiology* **16**(6), 593–605.
URL: <https://www.sciencedirect.com/science/article/pii/S074000209990285X>
- Kolassa, S. (2011), ‘Combining exponential smoothing forecasts using akaike weights’, *International Journal of Forecasting* **27**(2), 238–251.
URL: <https://www.sciencedirect.com/science/article/pii/S0169207010001032>
- López, S., Prieto, M., Dijkstra, J., Dhanoa, M. and France, J. (2004), ‘Statistical evaluation of mathematical models for microbial growth’, *International Journal of Food Microbiology* **96**(3), 289–300.
- Paine, C. E. T., Marthens, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A. and Turnbull, L. A. (2012), ‘How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists’, *Methods in Ecology and Evolution* **3**(2), 245–256.
URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2011.00155.x>
- Tsoularis, A. and Wallace, J. (2002), ‘Analysis of logistic growth models’, *Mathematical Biosciences* **179**(1), 21–55.
- Vithalani, P., Mahla, P., Padhiar, J., Bhanushali, U. and Bhatt, N. (2024), *Microbial Inventions and Improvement in Industrial Bioprocess Development*, Springer Nature Singapore, pp. 551–572.

- Váradi, L., Luo, J. L., Hibbs, D. E., Perry, J. D., Anderson, R. J., Oren, S. and Groundwater, P. W. (2017), ‘Methods for the detection and identification of pathogenic bacteria: past, present, and future’, *Chem. Soc. Rev.* **46**, 4818–4832.
- Wang, J. and Guo, X. (2024), ‘The gompertz model and its applications in microbial growth and bioproduction kinetics: Past, present and future’, *Biotechnology Advances* **72**(3), 08335.