

**Observing Microbial Population Growth across temperature
gradients: A Comparison of the performance of phenomenological
and mechanistic mathematical growth models**

Sebastian Dohne

CID: 01776389

Course Lead:
Saamrat Pawar

Word Count: 2767

A Mini-Project report submitted for the MSc in:
Computational Methods in Ecology and Evolution

Department of Life Sciences
Imperial College London
2025

Abstract

The mathematical modeling of microbial population has had an undeniable impact on diverse fields and industries, ranging from healthcare, biotechnology and agriculture and food, where phenomenological and mechanistic models are often employed to understand, and thus predict fundamental biological processes. In the context of Microbial Growth, I evaluated the performance of 3 models often used to predict growth, the cubic, logistic and gompertz models, in order to compare their robustness, flexibility and comparative accuracy towards the end of fitting to microbial growth data. Using a aggregated dataset consisting of 305 separate experiments collated from 10 studies, the models were fit using non linear least squares regression, a stochastic multi-starting parameter approach was initialized to optimize fitting. Models were compared using classical statistical techniques such as adjusted R^2 aswell as the information criterion AICc, comparing overall performance, and attempting to highlight differences in performance across temperature gradients. The study found the Gompertz model to statistically outperform the logistic and cubic models winning in all 3 main comparison categories (Adjusted R^2 , AICc and AICc weight (95%). The study however found that competitiveness increased with temperature. The study highlights the power of mechanistic model selection, in order to maximise parsimony, flexibility, robustness and biological interpretability.

1 Introduction

The historical use of mathematical modeling of bacteria has provided key insight for many fields and industries such as medicine, food safety, biotechnology and environmental science among many others. With Metabolic pathways, adaptability to extreme conditions, microbial interactions and the control of pathogenic bacteria are just some of the areas that the use phenomenological, empirical and mechanistic models have provided insight into. Microbial growth, is a fundamental field in microbiology and the employing these methods have allowed us to identify, define and predict its key stages [6]. The typical growth curve for bacteria in a closed system, manifests itself in an initial lag phase which can be considered an adaptation phase, and is characterized by cellular activity rather than growth where bacteria need to synthesis enzymes and coenzymes for growth, this then followed by exponential growth, slowing and plateauing into a stationary phase, where carrying capacity due to nutrient inavailability is reached, which is followed by a mortality phase as nutrient availability declines [12, 13, 16]. Temperature for instance, has a profound impact on the length of these phases as biochemical reaction rates increase with temperature according to the Boltzmann-Arrhenius equation towards an optimum, before thermal denaturation of enzymes occurs [5, 14, 27] accelerating growth.

Increasingly Model selection has proven to provide an alternative over the traditional approach of hypothesis testing, which focuses on rejecting arbitrary null hypotheses using arbitrary probability thresholds, as it directly compares multiple biologically meaningful hypotheses. Models are compared to another by evaluating relative support in the data, and are thus able to be ranked and weighted against another using tools such as information criteria [10]. using this framework, phenomenological models, which lack inherent biological meaning have been employed as benchmarks for comparison. In the context of bacterial growth modeling, such models like the linear polynomial cubic model have been employed for their ability to accommodate inflection points, allowing them to transition between growth phases, and while they have a risk of overfitting, it is less so than its higher-order counterparts [8]. With polynomial models like these however due to their phenomenological nature, it is difficult to ascertain biological significance from generated coefficients [15, 16].

Mechanistic models, which are derived from theoretical understanding of underlying processes are also employed, incorporating specific biological mechanisms with parameters having clear biological interpretation. The Logistic growth model based on the Verhulst model has been used to model bacterial growth and been modified in many different ways [16]. It is characterized by a sigmoidal growth curve, which resembles the lag, exponential and stationary phases closely, despite only having 3 parameters in its standard form, carrying capacity, intrinsic growth rate and initial population size. As such, it is a very parsimonious model and thus generalizable. Another empirical sigmoidal model is The Gompertz model, which is frequently employed model when observing bacterial growth specifically due to its flexibility with it being used particularly frequently when observing bacterial growth as its parameters can be adjusted to account for lag time specifically if required, unlike the standard logistic model. Due to their sigmoidal shapes however, both models suffer when accounting for the final bacterial stage in closed systems, where they fail to capture the mortality phase [4, 16].

The 3 models, Cubic, Logistic and a modified Gompertz function were fit to a bacterial dataset and compared using classical statistical approaches and information criteria, in an effort to observe how consistently each perform and how robust they are on a varied microbial species dataset, which were grown on a selection of media, at a variety of temperatures with different measurement units. Particular importance was placed on various temperature gradients to in an effort observe model responses to changes in bacterial growth patterns.

2 Materials and Methods

2.1 Data Collection

A large microbial dataset was aggregated from the findings of various studies totaling 305 separate experiments - with each being given a group ID - observing microbial growth data over time at various temperatures (ranging from 0-37°C) and was subsequently utilized for analysis [2, 3, 7, 9, 11, 17, 20–22, 27]. The dataset included a diverse range of bacterial genera from various taxonomic groups, including gram-positive and gram-negative organisms. Microbes from various environmental niches were represented, which included: common laboratory strains, environmental isolates, and clinically relevant species. Substrates on which these bacteria were grown in also varied from laboratory broths and agar as-well as meat and dairy based media. Growth was recorded in experiments over time in hours using either: OD-595, cell number, colony-forming units, and dry weight.

This data set was then refined, eliminating the collection errors that occurred during each experiment, such as recording values twice, removing microbial growth and time-point values that were negative and removing NA's in the data. Bacterial mass was plotted against time for each group, and groups with severely flawed data collection were manually excluded from the analysis. Finally, bacterial groups with less than data 6 points were excluded from the analysis, as it was deemed that meaningful models could not be fitted to the data. After cleaning the data, 274 of the 305 experiments remained. This subset of compiled growth formed the foundation for analysis in order to observe model effectiveness when observing bacterial growth data.

In order to compare differences between model accuracy across temperature gradients, these experiments were again split up into 3 groups, representing temperature gradients to be analysed, these were defined as low (0-12°C) consisting of 107 experiments, medium (12-25°C) with 88 experiments and high (25-37°C), 72 experiments. The temperature ranges were defined as such reflect cold, ambient temperatures and environments which facilitate optimal growth rates [19]. This allowed me to compare where models perform best when bacteria exhibit behaviors at different temperatures.

2.2 Models

To the growth dataset, the following phenomenological and mechanistic models were then fit:

The **linear cubic model** was fit using the standard formula:

$$f(x) = ax^3 + bx^2 + cx + d$$

The **non-linear logistic model** was fit using the formula [24, 25]:

$$N_t = \frac{N_0 K e^{rt}}{K + N_0(e^{rt} - 1)}$$

And the **non-linear modified Gompertz** was fit using the formula [26]:

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

Of which their parameters are:

- N_0 : population size for time 0
- N_t : population size for time t
- K : the carrying capacity
- r : growth rate
- N_{max} : maximum population size
- r_{max} : maximum growth rate
- t_{lag} : lag time before growth

2.3 Model Fitting and Comparisons

The cubic, Logistic and Gompertz models were each fitted to bacterial growth data using the Levenberg-Marquardt algorithm using Non-linear least squares regression. This was because algorithm lends itself well to these growth curves due to its ability to adjust its tuning parameter λ to switch between two strategies: gradient descent and Gauss-newton, allowing for fast and robust convergence on the correct coefficients even with noisy or sparse data, while minimizing the residual sum of squares. It also prevented parameters from becoming too large during fitting [23].

After the cubic function was fit to the data. Using nlsLM, for the Logistic(N) and Gompertz(Log(N)) models starting values were inputted for each each group-ID, and lower bounds were established. Starting values were defined as the following: For the logistic and Gompertz, N_0 was defined as the minimum value in growth data and K as 1.1x the maximum growth value, with r as 0.1. For the Gompertz, t_{lag} was estimated to be at the point between values where acceleration of growth-rate was greatest (r_{max}), the beginning of the exponential phase. Lower bounds were set to ensure biologically meaningful parameter estimates and to prevent fitting errors. To avoid 0 division errors, the growth rate was constrained to $r_{max} \geq 1 \times 10^{-5}$. The carrying capacity was restricted to $N_{max} \geq \min(\log N)$ in the Gompertz model and $K > 0.5 \times \max(PopBio)$ in the logistic model to ensuring valid population limits. The initial population size N_0 was -1 of the the smallest $\log N$ in the Gompertz model to accommodate log transformation but was constrained to $N_0 \geq 0$ in the logistic model. Finally, the lag time was set to $t_{lag} \geq 0$, as it could not be negative.

When the models were fit, I used a stochastic multi-start optimization approach. This was achieved by initiating 200 models to be fit for each group ID by drawing a value from a normal distribution with an SD of 0.1 for each parameter, while ensuring each parameter was at least 0. adjusted R^2 and AICc aswell as other metrics that could be used for comparison were stored when model fitting and the 200 models were then compared against another by using R squared and AICc values. The first 50 models of each run were compared by R squared value to establish a strong starting model, and the subsequent 150 models generated were then compared against another using their AICc values to produce an optimized model for logistic and gompertz per group-ID. The reason why AICc values were selected as the main comparator between models is because of its ability to balance model fit with complexity, penalizing overfitting while generalizing well. Specifically the corrected term was used over the standard AIC for its bias correction term for small sample sizes -which many of the group-IDS have- and it's ability to converge towards the AIC regardless when comparing n to the number of model parameters [1, 10].

As a result, optimal logistic and gompertz models were fit to each of the 274 groups. However, in less than 8 groups, convergence was not reached for either cubic, Gompertz or logistic and these were

removed for inter-model comparison. Relevant model metrics were then recorded. For each Group-ID where the phenomenological and mechanistic models were compared, frequencies were recorded for each group where $\Delta AICc > 2$ [1], adjusted R squared was the highest and Akaike weight was at least 0.95 for the model representing a 95% confidence interval [18] as they were defined as winning models in each of these categories. And observed across the low, medium and high temperature gradients.

2.4 Computing Tools

Analysis of the dataset was completed entirely in R Version - 4.3.3. The Tidyverse ecosystem was used for data wrangling using dplyr, as-well as plotting and visualization with GGPLOT2, GGPUBR. The Gridextra and Cowplot packages were also used for arranging and improving visualisation for generated plots. Model fitting using NLLS was performed using the Levenberg-Marquand algorithm via the use of the minipack.lm package. The parallel package was also used to maximize local computing speed when implementing the stochastic multi-start optimization approach for fitting the models.

3 Results

Overall, the model with the strongest support was found to be the Gompertz model as for all categories it had the highest win count, suggesting it is best supported by the data, showing substantially better trade-offs between fit and complexity for the most group-IDs. Consistency seems to diminish however when inspecting across temperature ranges, the logistic and cubic models became more competitive at with increasing temperatures.

Table 1: Frequency count of model “winners” across temperature ranges

Criterion	Model	Low(107)	Medium(88)	High(79)	Total(274)
Highest Adj-R ²	Cubic	20	18	29	67
	Logistic	21	28	11	60
	Gompertz	66	42	39	147
AIC Weight (95% CI)	Cubic	6	15	18	39
	Logistic	7	5	18	30
	Gompertz	89	63	32	184
AICc Winner $\Delta AICc > 2$	Cubic	7	15	23	45
	Logistic	9	7	21	37
	Gompertz	89	66	33	188

Note: Temperature ranges shown as Low (0-12°C), Medium (12-25°C), and High (25-37°C).

This trend continues when observing further model metrics, indicating that while Gompertz has a much lower mean AICc value at lower temperatures, this gap in model performance begins to narrow when temperatures begin to increase. When comparing the logistic against the cubic model, the cubic consistently outperformed the logistic in its adj-R² score and but the logistic had a lower AICc score for all categories.

Table 2: Model performance metrics across temperature ranges

Model	Adj-R ² [95% CI]	Mean AICc
Low Temperature (0-12°C)		
Cubic	0.895 [0.859–0.931]	71.67
Logistic	0.866 [0.799–0.933]	59.84
Gompertz	0.932 [0.890–0.975]	-36.52
Medium Temperature (12-25°C)		
Cubic	0.892 [0.869–0.914]	119.23
Logistic	0.854 [0.805–0.903]	106.79
Gompertz	0.943 [0.924–0.961]	-23.84
High Temperature (25-37°C)		
Cubic	0.813 [0.763–0.863]	17.66
Logistic	0.793 [0.734–0.852]	9.78
Gompertz	0.838 [0.780–0.897]	-35.51
Overall (All Temperatures)		
Cubic	0.870 [0.850–0.892]	71.37
Logistic	0.841 [0.806–0.876]	60.50
Gompertz	0.908 [0.883–0.933]	-32.17

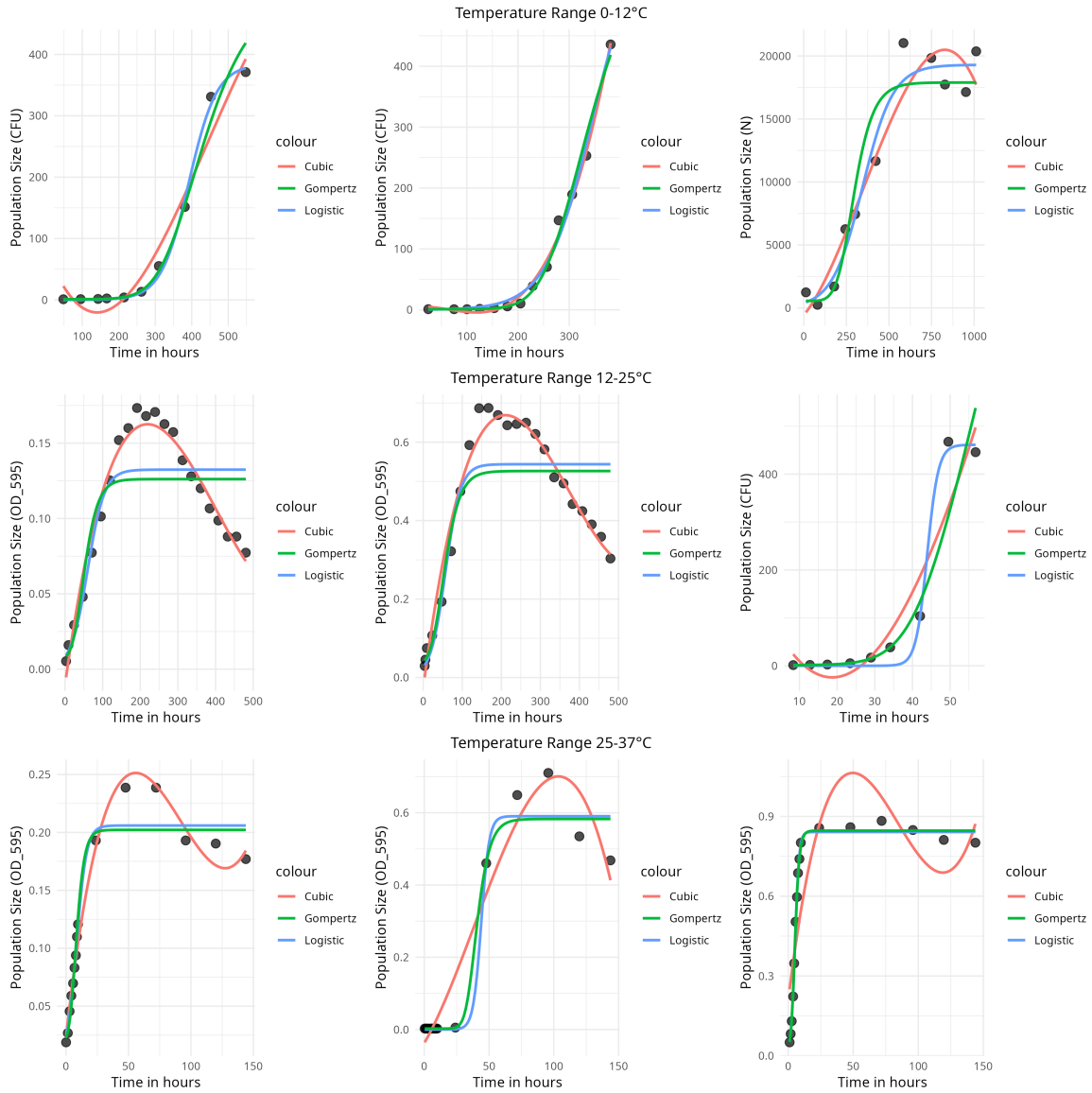


Figure 1: Example Group Subsets Showing Model Performance Across Different Temperature Ranges: All three models generally matched growth patterns. Distinct differences in model performance could be observed however at different temperatures. At low temperatures, all models fit relatively well, but the logistic and gompertz more accurately matched early growth patterns. At medium temperatures Gompertz and logistic were able to capture stationary phases more effectively where the cubic showed evidence of overfitting to the mortality phase in the data. At higher temperatures the cubic over and underestimated stationary growth where the logistic and gompertz formed a plateau.

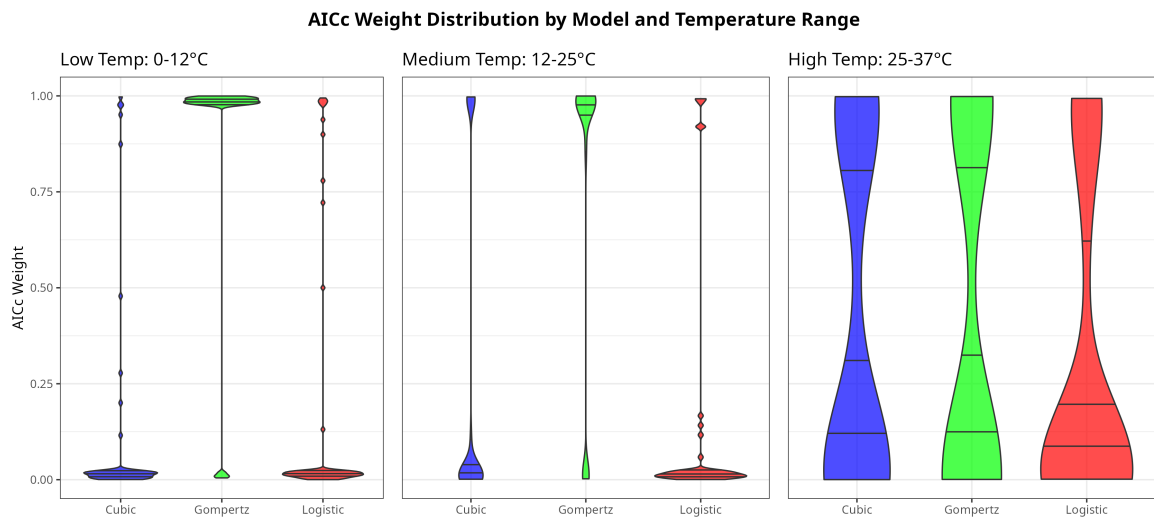


Figure 2: AICc Weight Distribution for Logistic, Gompertz, and Cubic Models for low, medium and high temperatures with lines in the violins representing 0.25, 0.5 and 0.75 probability quantiles. Gompertz outperforms Logistic and cubic in lower temperatures but it loses dominance at medium temperatures. At high temperatures a much more balanced distribution of AICc is observed indicating higher uncertainty and variability in model selection.

4 Discussion

Overwhelmingly, the results of this analysis have proven the Gompertz model to be the clear winner over the logistic function and the cubic model in its use-case in fitting to, and predicting bacterial growth. With this being especially evident at lower temperature ranges where it had almost 10x the winner score of the nearest model in terms of AICc and Aikake weight and was 3x more likely to have a higher adjusted R^2 value than the closest competing model. While the Gompertz "won" the most times in all categories with this dominance shifting towards optimum bacterial growth temperatures proven by the fairly even distribution of Aikake weights in this category. These results are perhaps unsurprising and have been supported empirically [4, 16, 26]. By design, the Gompertz allows for flexibility in its growth curve via the usage of its t_{lag} parameter, thus providing an advantage when it comes to capturing early growth patterns. When this factor becomes less of an advantage however, at it has been demonstrated with increased experimental temperatures, other models begin to perform comparatively.

The results of this experiments have also ended up supporting the previous understanding of each model, with the focus of temperature gradients highlighting the features of each one with their use case. Despite Gompertz being the clear winner according to the information criteria measured by a wide margin. All models have proved themselves to fit well to the data, despite experimental groups having a plethora of different starting conditions from measurement units, media, species measured and temperature, having very high adj R^2 scores demonstrating their use case in the context of estimating and predicting bacterial growth with how generalized they are. Calculating information criteria like AICc and its weights has however allowed us to understand why the Gompertz model definitely is and the logistic model may still be a better choice than linear polynomial functions.

It was visually observed that both the logistic and Gompertz were able to more effectively capture lag phases and stationary phases across group-IDs due to their sigmoidal shapes, with their flexibility allowing them to fit well to at increased temperatures, where exponential growth occurred sooner giving them more biological relevance, which is further reflected that even in cases where the cubic may fit the best out of the non-linear functions, it is difficult to ascertain any biological significance out of its coefficients, which isn't to say it hasn't been experimentally attempted [8]. It is worth raising that in cases where the cubic did outperform the mechanistic models, it may have been attributed to the event that the other two mechanistic models fail to consider, namely the death phase [16], since it is usually observed closed systems. Where the two models plateau towards the maximum observed bacteria per group, the cubic was observed using its inflection points to overfit to the declining bacterial data, potentially increasing its perceived efficacy in modeling bacterial behavior, highlighting the importance of information criteria to spot actions like this [10].

Despite the systematic approach to model selection this study utilized. It presents several areas of improvement which should be considered when interpreting the results. While the heterogeneity of the dataset provided a strong vehicle to demonstrate the Generalisability of the models. It may have diminished the ability of the models to detect underlying patterns in the data, with frequent temporal variability across all group-IDs in measurement providing potential issues for models to accurately capture the complete growth curve. Also while in total, the data shows strong support for the Gompertz with 274 groups being compared, with the lower experiment numbers for each temperature gradient, this general assessment can be scrutinized. Finally, Model selection itself could have been refined further, less parsimonious primary models, like the Baranyi model could have been considered, which is specific to microbial growth, incorporating a lag phase in its parameters over a singular lag time, as well. The utilization of splines in the form of three phase linear methods also could add value to the experimental method as it has done in past research [4].

5 Conclusions

I successfully fit and compared the phenomenological and mechanistic cubic, logistic and Gompertz functions using NLLS and information criteria to provide insight into their comparative overall robustness, as well as their performance across temperature gradients on a highly varied microbial dataset. While all models fit well to the data proving generalising abilities. The results supported established the wide usage of the Gompertz in the literature in regards to bacterial growth modeling, as it outperformed both the non-linear logistic equation and the cubic equation across all temperature estimates according to all comparison metrics utilized, waning dominance however when approaching higher temperatures. In model strength and weaknesses were identified and confirmed, providing further supporting evidence in supported ideas in the field of microbial modelling.

References

- [1] D Anderson and K Burnham. Model selection and multi-model inference. *Second*. NY: Springer-Verlag, 63(2020):10, 2004.
- [2] Young-Min Bae, Ling Zheng, Jeong-Eun Hyun, Kyu-Seok Jung, Sunggi Heu, and Sun-Young Lee. Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce. *Journal of Food Science*, 79(10):M2072–M2080, 2014. Publisher: Wiley Online Library.
- [3] Joey R Bernhardt, Jennifer M Sunday, and Mary I O’Connor. Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity. *The American Naturalist*, 192(6):687–697, 2018. Publisher: University of Chicago Press Chicago, IL.
- [4] RL Buchanan, RC Whiting, and WC Damert. 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, 1997. Publisher: Elsevier.
- [5] Anthony I Dell, Samraat Pawar, and Van M Savage. Systematic variation in the temperature dependence of physiological and ecological traits. *Proceedings of the National Academy of Sciences*, 108(26):10591–10596, 2011. Publisher: National Academy of Sciences.
- [6] Daniel S Esser, Johan HJ Leveau, and Katrin M Meyer. Modeling microbial growth and dynamics. *Applied microbiology and biotechnology*, 99:8831–8846, 2015. Publisher: Springer.
- [7] Liane Aldrighi Galarz, Gustavo Graciano Fonseca, and Carlos Prentice. Predicting bacterial growth in raw, salted, and cooked chicken breast fillets during storage. *Food Science and Technology International*, 22(6):461–474, 2016. Publisher: SAGE Publications Sage UK: London, England.
- [8] Blanca E Garcia, Emmanuel Rodriguez, Yolocauhtli Salazar, Paul A Valle, Adriana C Flores-Gallegos, O Miriam Rutiaga-Quñones, and Raul Rodriguez-Herrera. Primary model for biomass growth prediction in batch fermentation. *Symmetry*, 13(08):1468, 2021. Publisher: MDPI.
- [9] CO Gill and KM DeLacy. 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, 1991. Publisher: Elsevier.
- [10] Jerald B Johnson and Kristian S Omland. Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2):101–108, 2004. Publisher: Elsevier.
- [11] S Rohini Karnat, MVRK Sarma, and D Somashekar. Study of changes in functional properties of fermented Idli batter during storage under different temperature conditions. *Food and Humanity*, 3:100377, 2024. Publisher: Elsevier.
- [12] Anna Lo Grasso, Ada Fort, Fariba Fahmideh Mahdizadeh, Agnese Magnani, and Chiara Moccenni. Generalized logistic model of bacterial growth. *Mathematical and Computer Modelling of Dynamical Systems*, 29(1):169–185, 2023. Publisher: Taylor & Francis.
- [13] Raina M Maier and Ian L Pepper. Bacterial growth. In *Environmental microbiology*, pages 37–56. Elsevier, 2015.
- [14] Leonor Michaelis, Maud L Menten, and others. Die kinetik der invertinwirkung. *Biochem. z*, 49(333-369):352, 1913. Publisher: Berlin.
- [15] Harvey Motulsky and Arthur Christopoulos. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press, 2004.

- [16] Micha Peleg and Maria G. Corradini. Microbial growth curves: What the models tell us and what they cannot. *Critical Reviews in Food Science and Nutrition*, 51(10):917–945, 2011. Publisher: Taylor & Francis tex.eprint: <https://doi.org/10.1080/10408398.2011.570463>.
- [17] JD Phillips and MW Griffiths. The relation between temperature and growth of bacteria in dairy products. *Food Microbiology*, 4(2):173–185, 1987. Publisher: Elsevier.
- [18] David Posada and Thomas R Buckley. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5):793–808, 2004. Publisher: Society of Systematic Zoology.
- [19] David A Ratkowsky, June Olley, TA McMeekin, and Andrew Ball. Relationship between temperature and growth rate of bacterial cultures. *Journal of bacteriology*, 149(1):1–5, 1982.
- [20] Norman G Roth and Robert B Wheaton. Continuity of psychrophilic and mesophilic growth characteristics in the genus *Arthrobacter*. *Journal of Bacteriology*, 83(3):551–555, 1962.
- [21] Kaarina Sivonen. 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, 1990.
- [22] CJ Stannard, AP Williams, and PA Gibbs. Temperature/growth relationships for psychrotrophic food-spoilage bacteria. *Food Microbiology*, 2(2):115–122, 1985. Publisher: Elsevier.
- [23] Mark K Transtrum and James P Sethna. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. *arXiv preprint arXiv:1201.5885*, 2012.
- [24] Anastasios Tsoularis and James Wallace. Analysis of logistic growth models. *Mathematical biosciences*, 179(1):21–55, 2002. Publisher: Elsevier.
- [25] Pierre-François Verhulst. Notice sur la loi que la population suit dans son accroissement. *Correspondence mathématique et physique*, 10:113–129, 1838.
- [26] Marcel H Zwietering, Ida Jongenburger, Frank M Rombouts, and KJAEM Van’t Riet. Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–1881, 1990.
- [27] MH Zwietering, JC De Wit, HGAM Cuppers, and K Van’t Riet. Modeling of bacterial growth with shifts in temperature. *Applied and Environmental Microbiology*, 60(1):204–213, 1994.