

# Biphasic Logistic Models Outperform Simpler Ones in Describing Bacterial Growth Curves

Harry Trevelyan

March 2025

## Abstract

This study investigates the relative merits of the logistic, Gompertz, and biphasic logistic models of population growth when applied to a large dataset of bacterial growth curves from multiple species. These models were fit using non-linear least squares minimization, where I optimized the model parameters to best fit the observed data for each bacterial time series. The logistic model was the most widely applicable to bacterial population curves although relatively difficult to fit and interpret, the Gompertz model is often appropriate but fails to capture population decline, and the biphasic logistic model provides an intuitive description of microbial growth in idealised growth scenarios without death or habituation. The results of this study provide insight into the trade-offs between goodness-of-fit, parsimony, and applicability.

## 1 Introduction

Models in science aim to capture the essential features of phenomena. They have many uses: their underlying assumptions may form testable hypotheses; their assumptions may be adjusted to predict a system's behavior in different conditions; and they may tell us about how different factors interact within complex systems.

However, models must be designed to suit the specific needs and goals of their users. Levins [1] considered three ways to evaluate a model: how accurately it describes the system it was made for (realism), its applicability to other systems (generality), and how precisely it makes predictions or quantifies system behavior (precision). Another important aspect is interpretability- how easily a model's insights can be understood and applied within its context. Related to this concept is the principle of parsimony, which states that the best model is the simplest one that can successfully predict, explain or describe a phenomenon. There are also practical considerations to model selection: a powerful model whose inputs are impossible to obtain will have limited applicability in practice. These six features- realism, generality, precision, interpretability, parsimony, and ease of use- make a useful framework for evaluating how successfully models meet their goals. Obviously, no model can simultaneously maximise all these desirable attributes [1]. Thus, different fields of science have developed specific strategies of model-building based on their unique challenges and aims.

Microbial population growth is an area of biology that has been the subject of extensive modeling. Understanding how these populations grow and decline is of obvious interest to fields such as food microbiology, epidemiology, biotechnology, and climate science, as fluctuations in microbial density can significantly impact key emergent properties like carbon cycling, food spoilage, and disease spread. Because microbial growth is influenced by numerous variables, it is essential to construct models that account for these complexities while remaining practical and interpretable. Microbiologists must decide on the appropriate level of simplicity, specificity, accuracy, precision, practicality and interpretability to build into models of growth.

Many models are available to describe the growth of microbial populations. Among the simplest is the logistic growth equation, which predicts the growth pattern of a population from its starting size, growth rate, and theoretical maximum population size or carrying capacity [2] (equation 1, figure 1a). However, the logistic model makes several simplifying assumptions—it predicts the dynamics for a single, well-adapted population in an unchanging environment with no death. Other models of population growth aim to enhance their generality and realism by relaxing one or more of these assumptions: the Gompertz model, for example, incorporates a fourth 'time lag' parameter (equation 2, figure 1b, [3]) to account for scenarios where initially poorly adapted populations experience a delay for habituation before growth accelerates. Real microbial populations typically decline after reaching their maximum size due to a buildup of toxic waste products and/or a depletion of resources [4]. A biphasic logistic model can capture this death phase using five parameters [5]: the initial three parameters of the logistic equation, a term describing the onset of the death phase, and the death rate (equation 3, figure 1c).

In this study, I will investigate the relative merits of the logistic, Gompertz, and biphasic logistic models of population growth when applied to a large dataset of bacterial growth curves from multiple species. I will evaluate the models based on their accuracy, precision and generality, as well as their ease of use, parsimony and interpretability. Given the varying complexity of these models, I hypothesize that while the logistic model will perform well in simple, ideal growth scenarios, the Gompertz and biphasic logistic models will provide more accurate predictions in cases with delayed growth phases and population decline. The results of this study will inform the selection of appropriate models for microbial growth in various industrial, medical, and ecological applications, and provide insight into the trade-offs involved in model selection.

#### Model equations:

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

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

$$N(t) = \frac{K}{1 + \left( \frac{K - N_0}{N_0} \right) e^{-rt}} - D e^{-\mu t} \quad (3)$$

#### Symbol Key:

- $N(t)$ : Population size at time  $t$
- $N_0$ : Initial population size
- $K$ : Carrying capacity
- $r$ : Maximum growth rate
- $t$ : Time
- $t_{\text{lag}}$ : Time lag (specific to Gompertz model)
- $D$ : Death rate parameter (specific to Biphasic Logistic model)
- $\mu$ : Decay rate parameter (specific to Biphasic Logistic model)

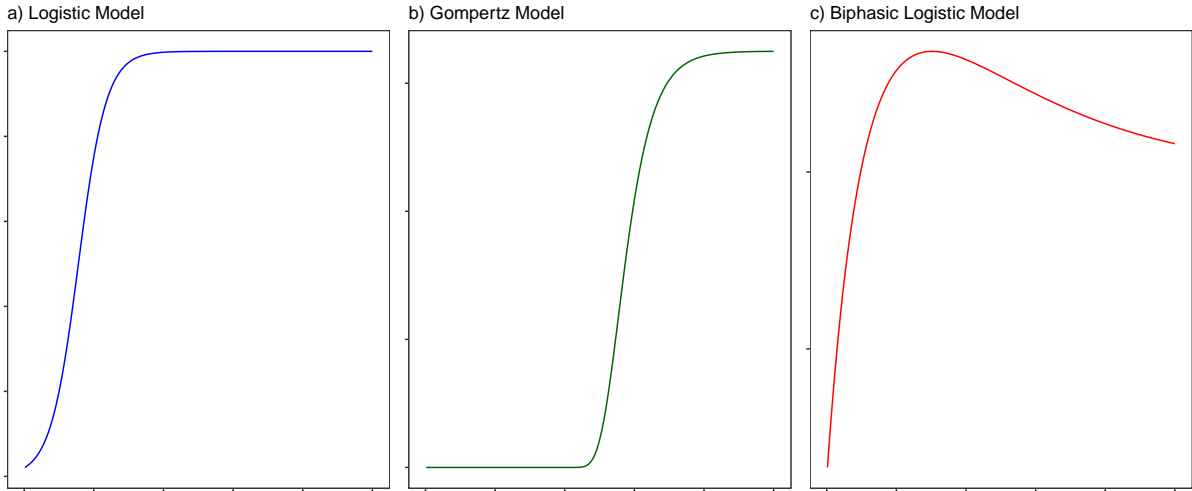


Figure 1: Population Growth Models: Logistic, Gompertz, and Biphasic Logistic. Abundance is plotted against time. a) In the logistic model, a population grows exponentially, and growth slows to zero as it approaches its carrying capacity. b) The Gompertz model displays the same pattern but incorporates a lag phase before growth. c) In the biphasic logistic model, populations decline after reaching their peak.

## 2 Methods

### 2.1 Raw data and transformations

My dataset consists of 305 microbial biomass or cell count measurements over time from 10 studies worldwide [6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. These time series were narrowed to 179 after excluding those with fewer than 10 data points or with clear errors in data collection (12 erroneous time series were identified by visual inspection of growth curves). The studies used four methods to measure abundance: optical density at 595 nm, cell count, colony-forming units, and dry weight. Differences in units were assumed to have no impact on model fitting.

Abundance data were log-transformed for three reasons: to address right skewness in the data, to improve the normality of residuals in non-linear models, and to ensure consistent data to allow model comparison. Although the logistic and biphasic logistic models are typically run on raw abundance data, I found that log transformation actually improved the homoscedasticity of the residuals in these models and improved the ease of model fitting.

### 2.2 Model fitting

Six models were fitted to the log-abundance time series data: the logistic, Gompertz, and biphasic logistic non-linear models, along with the first three orders of polynomials as linear models for comparison. Akaike Information Criteria were calculated to evaluate each model on each time series, correcting for small sample sizes (AICc). AICc rewards goodness of fit but penalises models with high numbers of parameters, thus balancing goodness of fit with parsimony.

Non-linear models require starting values for parameters, which are adjusted through fitting algorithms to minimize non-linear least squares (NLLS). Reasonable starting values were estimated for each time series using the approximations outlined in Table 1. To ensure optimal fitting, each non-linear model was fit 50 times per time series, with starting parameter values drawn from normal distributions around these approximations. The model with the lowest AICc value was selected for analysis. The sensitivity of each model to starting parameters was also recorded, calculated as the standard deviation of the AICc values from multiple fits on each time series. This sensitivity was used as a measure of model robustness.

Table 1: Parameter initialisation methods for each model.

Parameter	Model used in	How starting value was estimated
$N_0$	Logistic, Gompertz and Biphaseic	Minimum abundance in time series
$K$	Logistic, Gompertz and Biphaseic	Maximum abundance in time series
$r$	Logistic, Gompertz and Biphaseic	Maximum growth rate in time series
$t_{\text{lag}}$	Gompertz	Time at which second-order differential is maximum
$D$	Biphaseic	0.5 for all time series
$\mu$	Biphaseic	0.01 for all time series

### 2.3 Results analysis and visualisation

The data were subset to only those in which all six models had run successfully. For each time series, the model with the lowest AICc was identified and recorded. These results were summarised over all datasets to show how often each of the six model types was the most suitable for each time series. Parameter estimates were extracted from each model and used to generate a table of fitted values, which were plotted over the original data.

### 2.4 Computing languages and tools

Data wrangling, model fitting and visualisation were all completed in R (4.4.3). Linear models were fit with the `lm()` function from base R. Non-linear models were fit using the `nlsLM()` function from the `minipack.lm` package. This function uses the Levenberg-Marquardt algorithm to minimise NLLS. The `ggplot2` package was used for plotting.

## 3 Results

Of the 179 time series in my analysis, all six models successfully fit in 154 cases, with the biphaseic logistic model failing most frequently- likely due to poor initial parameter estimates. Despite this, the AICc of the logistic and Gompertz models were the most sensitive to starting parameters, showing higher standard deviation in AICc scores on models run on the same time series (Table 2). These results suggest that the logistic and Gompertz models can successfully run under a wider range of starting parameters than the biphaseic model, but once they do run, differences in these parameters can affect goodness of fit. In contrast, the biphaseic logistic model appears to converge more stably to an optimal solution when adequately initialised, but is more prone to fail given poor initial parameter estimates.

The logistic, Gompertz and biphaseic logistic models outperformed the polynomial models in almost all cases, with the biphaseic logistic model emerging as the best model most frequently (37%) (Table 2). There was little difference in the success rates of the logistic and Gompertz models (29.2% vs 26.6% respectively). The quadratic model had a higher success rate (3.9%) than the linear and cubic polynomials, which only outperformed the others in 1.9 and 1.3% of cases respectively. Plotting the fitted values alongside the original data highlights under which each non-linear model excels (figure 2 a,b,c): the biphaseic logistic model fits best when the time series data exhibit a death phase; the Gompertz model is optimal for growth curves with a clear time lag; and the logistic model is sufficient when neither lag nor death phase is present. These plots also show that polynomials can outperform the non-linear models by capturing the essential trends in the data with fewer parameters (figure 2 d, e, f).

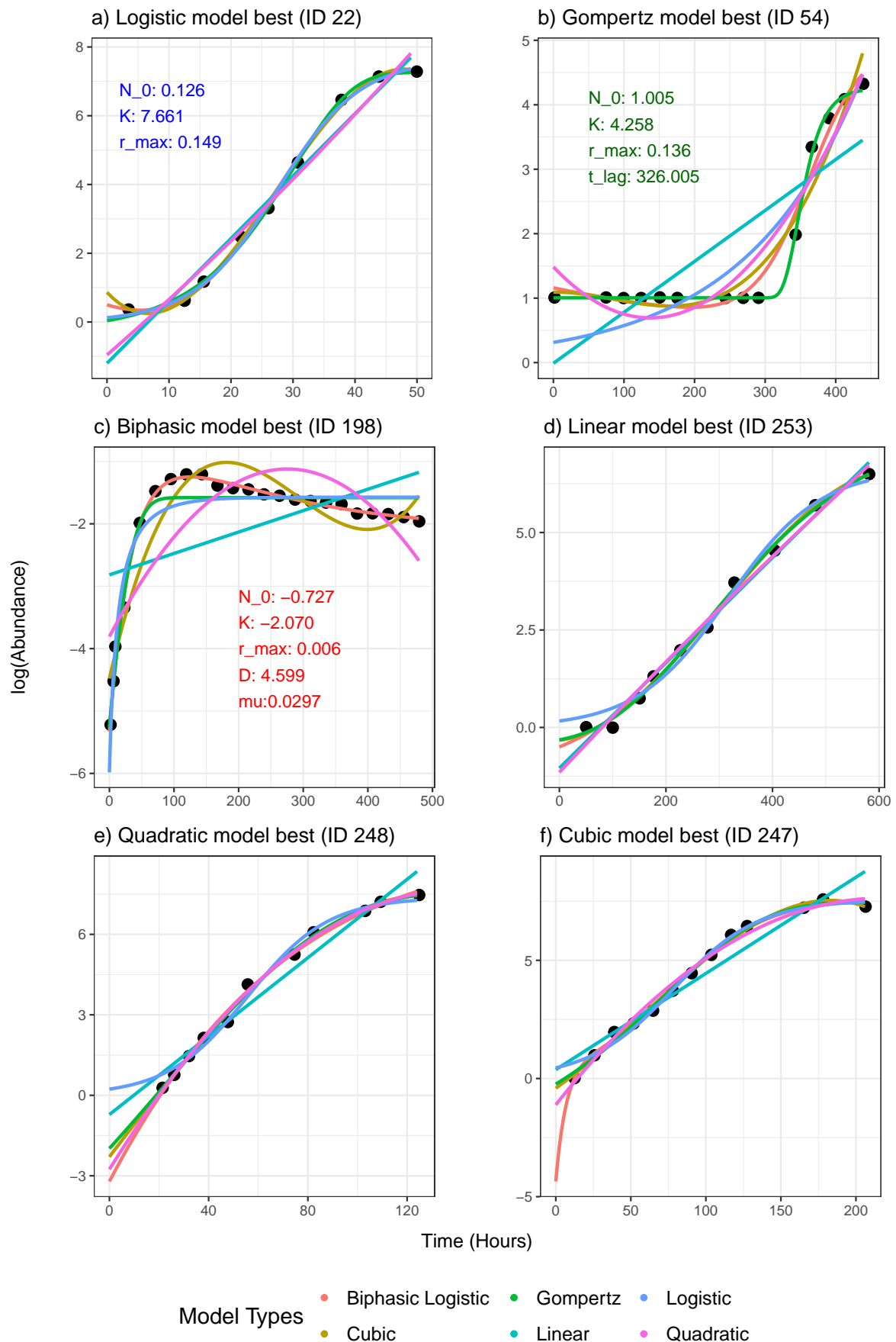


Figure 2: Model Fit Comparison. Six time series of bacterial abundance (black points) are shown, each with the six chosen models (logistic, Gompertz, biphasic, linear, quadratic, and cubic- coloured lines) overlaid. Above each subplot, the best model to describe those data (i.e. the model with the lowest AICc) is indicated. In plots a-c, parameter estimates of the best model are displayed. a) The logistic model performed best on datasets with exhibiting ideal growth conditions. b) The Gompertz model performed best on datasets showing a lag phase. c) The biphasic logistic model performed best on datasets with a death phase, but note its unintuitive parameter estimates. d-f) Polynomials outperformed non-linear models in time series when they could capture the patterns in data with fewer parameters.

Table 2: Model Performance Summary. In the 154 time series to which all 6 models could be fitted, the number of times each model had the lowest AICc (i.e. was the "winner") was counted and the win percentage calculated. For the non-linear models, the mean standard deviation of the AICc values from multiple fits on each time series is shown as a measure of model sensitivity to initial parameters.

Model	Win Count	Win Percentage (%)	AICc standard deviation
Biphasic Logistic	57	37	12.4
Logistic	45	29.2	16.9
Gompertz	41	26.6	16.5
Quadratic	6	3.9	N/A
Linear	3	1.9	N/A
Cubic	2	1.3	N/A

## 4 Discussion

Comparison of AICc scores across models revealed that the biphasic logistic model was most often the best at describing the data, even taking into account its relative complexity. This speaks to the accuracy and precision of the biphasic model over a wider range of data- that is, its high generality. However, this lead was only marginal (biphasic logistic model had a 37% win rate, compared to 29% for logistic and 27% for Gompertz- Table 2) and my results would benefit from bootstrapping analysis to calculate confidence intervals around these estimates. As hypothesized, the examples in Figure 2 show that each non-linear model outperformed the others when applied to data that best fit their respective equations. This suggests that a different ranking might emerge with a different collection of time series data. Regardless, the death phase is a genuine characteristic of bacterial growth curves, and it is well understood to be caused by nutrient depletion or toxin buildup [4]. This phenomenon is realistically and precisely captured by the biphasic logistic model. That a simple linear model could ever outperform more complex ones in terms of AICc (as it did 1.9% of the time) shows that the additional complexity introduced by more parameters may not significantly improve model fit and could even lead to overfitting, especially when the underlying data do not reflect such complexity. These results demonstrate that the enhanced realism and precision conferred by complex models is often not necessary, and highlight the importance of identifying patterns in data before model fitting.

As stated in my introduction, model evaluation encompasses more than just traditional fit metrics: how did the non-linear models perform in terms of ease of use and interpretability? The biphasic logistic model proved challenging to run: this model failed in 25 out of 179 runs, primarily because of the sensitivity of initial parameter estimates. However, once the model was able to converge, it did so with relative stability (Table 2), suggesting that with proper initialization, it could fit to all 179 datasets. The challenges in finding suitable initial values for the model parameters could potentially be addressed with more advanced techniques. For example, rather than arbitrary starting points, machine learning algorithms could optimize the initialization process. Additionally, while NLLS optimization provides only point estimates for the parameters, Bayesian methods could be used to quantify the uncertainty around these estimates. To illustrate the difficulty in initialising parameters for the biphasic logistic model, consider the parameter estimates of the optimal model in Figure 2c: the optimal solution included an  $N_0$  term of 0.006, a  $K$  term of -2.07, and a  $D$  term of 2.60. Clearly, while the biphasic model in Figure 2c fits the data well, the interaction between parameters leads to unintuitive estimates that are impossible to interpret biologically. In contrast, the logistic model's simpler, more mechanistic nature avoids these issues (Figure 2a). The logistic model's parameters are intuitively understandable and generate meaningful hypotheses, which are clear advantages over the biphasic and polynomial models. The Gompertz model had an intermediate level of ease of use and interpretability. While it occasionally produced biologically implausible estimates for parameters (such as negative time lags), its initial parameters could be readily estimated from the data and its outputs made sense of.

What model, then, to use? The Gompertz model is most widely used in the food microbiological literature [16], which seems to contradict my main result that the biphasic logistic is most widely applicable to bacterial growth curves. The contradiction is resolved upon consideration of the priorities of food microbiologists- this field is mostly concerned with describing and predicting the period before microbe populations have exploded, as after this food will have spoiled ([4], but see [17]). This example demonstrates how the trade-offs in model fitting are navigated: increasing realism at the expense of ease of use and interpretability is not valuable if the benefits do not apply to the focus of interest. The

biphasic logistic model I used in this study has its origins in pharmacology: like populations that boom and wane with time, drugs show analagous increases and decreases in efficacy as their dose increases, and the decline phase is of direct interest to modelers [5].

In conclusion, a model-maker must navigate trade-offs between goodness-of-fit, parsimony, and applicability. Scientists weigh these considerations depending on their specific research goals: for instance, while an economist might prioritize accuracy and precision in demographic projections, a conservationist might be more concerned with making qualitative predictions under varying scenarios (e.g. of global warming). In this exercise, I showed that the biphasic model is most widely applicable to bacterial population curves although relatively difficult to fit and interpret, the Gompertz model is often appropriate but fails to capture population decline, and the logistic model provides an intuitive description of microbial growth in idealised growth scenarios without death or habituation. I close by emphasising that model choice is guided not just by the data, but by the questions being asked and the practical utility the model provides in answering those questions.

## References

- [1] R. Levins. The strategy of model building in population biology. *Am. Sci.*, 54:421–431, 1966.
- [2] David M Bradley. Verhulst’s logistic curve. *The College Mathematics Journal*, 32(2):94–98, 2001.
- [3] M. H. Zwietering et al. Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6):1875–1881, 1990.
- [4] P. Micha and M. G. Corradini. Microbial growth curves: What the models tell us and what they cannot. *Critical Reviews in Food Science and Nutrition*, 2011.
- [5] W. N. Beckon et al. A general approach to modeling biphasic relationships. *Environmental Science & Technology*, 42(4):1308–1314, 2008.
- [6] Y. M. Bae et al. Growth characteristics and biofilm formation of various spoilage bacteria isolated from fresh produce. *Journal of food science*, 79(10):M2072–M2080, 2014.
- [7] J. R. Bernhardt et al. Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity. *The American Naturalist*, 192(6):687–697, 2018.
- [8] L. A. Galarz et al. Predicting bacterial growth in raw, salted, and cooked chicken breast fillets during storage. *Food Science and Technology International*, 22(6):461–474, 2016.
- [9] C. O. Gill and K. M. 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.
- [10] J. D. Phillips and M. W. Griffiths. The relation between temperature and growth of bacteria in dairy products. *Food Microbiology*, 4(2):173–185, 1987.
- [11] N. G. Roth and R. B. Wheaton. Continuity of psychophilic and mesophilic growth characteristics in the genus arthrobacter. *Journal of bacteriology*, 83(3):551–555, 1962.
- [12] A. P. R. D. Silva et al. Modeling the growth of lactic acid bacteria at different temperatures. *Brazilian Archives of Biology and Technology*, 61, 2018.
- [13] K. Sivonen. Effects of light, temperature, nitrate, orthophosphate, and bacteria on growth of and hepatotoxin production by oscillatoria agardhii strains. *Appl. Environ. Microbiol.*, 56(9):2658–2666, 1990.
- [14] C. J. Stannard et al. Temperature/growth relationships for psychrotrophic food-spoilage bacteria. *Food Microbiology*, 2(2):115–122, 1985.
- [15] M. H. Zwietering et al. Modeling of bacterial growth with shifts in temperature. *Appl. Environ. Microbiol.*, 60(1):204–213, 1994.
- [16] József Baranyi and Terry A Roberts. A dynamic approach to predicting bacterial growth in food. *International journal of food microbiology*, 23(3-4):277–294, 1994.

- [17] J. Horowitz et al. Probabilistic model of microbial cell growth, division, and mortality. *Appl Environ Microbiol*, 76:1057–1064, 2010.