# A Model Comparison Approach for Bacterial Growth: Quadratic, Logistic, and Gompertz Models with and without Log Transformation

**Anna Cavalieri Canosa**
CID 06004153

MSc Computational Methods in Ecology and Evolution
Department of Life Sciences
Imperial College London

Word Count: 2348 words

March 14, 2025

# Abstract

Mathematical modelling is crucial in understanding bacterial growth dynamics, offering predictive insights relevant to healthcare, food safety and industrial microbiology. Despite its importance and common use, selecting an appropriate bacterial population growth model remains a challenge due to the complexity and variability of bacterial population behaviours. In this study, I compare Quadratic, Logistic and Gompertz models, assessing their ability to capture bacterial growth phases using both untransformed raw and log-transformed population data. Model performance was evaluated using $R^2$, AIC, AICc, and BIC, with the better model determined by the highest $R^2$ and lowest information criteria values. Results indicate that the log Gompertz model consistently provided the overall best fit particularly excelling in capturing lag phase due to its additional $t_{\text{lag}}$ parameter which accounts for bacterial adaption time to new environments. However, the Log Logistic model outperformed in AICc, and none of the models accurately captured the death phase. The findings highlight the trade-off between model complexity and biological interpretability - suggesting that no single model universally fits all bacterial growth scenarios. Future research should integrate phase-specific models and environmental factors to improve predictive accuracy, particularly for applications in antibiotic treatment optimisation, microbial ecology, and industrial fermentation.

# Introduction

Explaining complex biological mechanisms is challenging, and predicting them seems almost unattainable. However, quantitative modelling significantly aids in demystifying the intricacies of biological phenomena and enhances predictive capabilities. A prime example is modelling bacterial population growth, which combines biological insights with mathematical precision. These models play crucial roles in a variety of sectors, including healthcare, the food industry, and environmental management. They assist in forecasting phenomena like the shelf life of perishables (Koseki 2016), optimizing antibiotics efficiency (Coates et al. 2018), and controlling the spread of infectious diseases like cholera (Singh et al. 2024; Brhane et al. 2024).

Despite its value, modelling bacterial growth is complex due to the dynamic growth patterns of bacteria. These patterns typically include an initial lag phase where bacteria acclimate to new conditions, followed by a rapid exponential growth phase, a stationary phase as resource limits are reached, and finally a decline or death phase (Kumakura et al. 2023; Bate et al. 2023; Monod 1949; Rolfe et al. 2012). Each of these phases represents different biological processes and challenges for modelling. This is possibly the reason why literature has yet to standardize a model for bacterial growth, although models like logistic, Gompertz, Baranyi, and Richards are commonly used (Jewell 2012; Huang 2013; López et al. 2004; Buchanan et al. 1997; Pla et al. 2015).

In my analysis, I compare three of these models - Quadratic, Logistic, and Gompertz — using both natural and log-transformed data.

- I selected to test the **Quadratic Model** for its simple linear form, which visually resembles the general pattern of bacterial growth, though I hypothesised it would struggle with accurately capturing the nuances of the lag and death phases due to its simplicity.

- I chose the **Logistic Model** as it is known for its ability to mirror growth as it approaches a carrying capacity, aligning well with bacterial behaviour under resource constraints, and I expect it to perform well during the stationary phase.

- While I selected the **Gompertz model** due to its time-lag parameter, enhancing model complexity and potentially providing a more accurate description of the initial adaptation period.

I hypothesise that while none of these models will perfectly capture the complete dynamics of bacterial population growth due to the diverse nature of the data, the Gompertz model will likely be the most effective of the three. Additionally, it's uncertain whether log transformation of the data will significantly impact the results, but it's anticipated that plotting logarithmic population biomass might better constrain variance, enhancing the model fit.

## Data Collection and Analysis

### Data Collection

Data were collected from 10 research articles: Bae et al. 2014; Bernhardt et al. 2018; Galarz et al. 2016; Gill et al. 1991; Phillips et al. 1987; Roth et al. 1962; Silva et al. 2018; Sivonen 1990; Stannard et al. 1985; Zwietering et al. 1994. The collected data yielded 305 datasets, encapsulating a diverse range of bacterial species pertinent to food spoilage and industrial fermentation, including *Lactobacillus*, *Pseudomonas*, and *Escherichia*. Experimental conditions varied extensively, with temperatures ranging from 0°C to 45°C, reflecting different storage and processing environments. Growth mediums included nutrient broths, dairy, and meat substrates, enhancing the ecological validity of the models. Additionally, some studies modified pH and salinity, further broadening the environmental scope assessed in relation to bacterial growth dynamics.

### Models

The Quadratic Model was fitted using formula:

$$P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 \tag{1}$$

The coefficients $\beta_0$, $\beta_1$, and $\beta_2$ are used for fitting the curve to the data, without directly representing biological processes.

The Logistic Model was fitted with (Pla et al. 2015):

$$P(t) = \frac{N_0 \cdot K \cdot e^{r \cdot t}}{K + N_0 \cdot (e^{r \cdot t} - 1)} \tag{2}$$

The Gompertz Model modification to fit modern statistical software was fit with equation (Gompertz 1825; Winsor 1932; Pla et al. 2015; Buchanan et al. 1997):

$$P(t) = N_0 + (K - N_0) \cdot \exp\left(-\exp\left(\frac{r \cdot e \cdot (t_{\text{lag}} - t)}{(K - N_0) \cdot \log(10)} + 1\right)\right) \tag{3}$$

Table 1: Overview of the parameters and the methodologies employed for initial estimates of the Logistic and Gompertz models. These estimates were carried out on the natural population dataset (PopBio) and on the logarithmic transformation(logPopBio).

| Parameter | Description | Estimation Method |
|---|---|---|
| $N_0$ | Initial population size | First observed value of the population in the dataset. |
| $K$ | Carrying Capacity | Maximum observed value of the population. |
| $r_{\max}$ | Maximum growth rate | Maximum value of the first-order difference of population size. |
| $t_{\text{lag}}$ | Lag time before exponential growth | Time at which growth rate reaches its maximum. |

The parameters used in both the Logistic and Gompertz models are biologically significant, representing key aspects of population growth dynamics. The initial estimates for these parameters were derived using methods outlined in Table 1, ensuring biologically meaningful starting values. These estimates were applied to both the natural population dataset (PopBio) and its logarithmic transformation (logPopBio) to enhance model robustness. The log-transformation was chosen to normalise distributions and reduce skewness. These values are also used to compare the effectiveness of linear and logarithmic representations in capturing the dynamics of bacterial population growth.

## Model Fitting and Evaluation

Following the estimation of biological meaningful initial parameters Table 1, the Logistic and Gompertz models were fitted to individual bacterial growth data subsets using the Levenberg-Marquardt algorithm for non-linear least squares estimation. This approach was chosen due to its efficiency in converging to a global optima and avoiding local minima, essential for accurate parameter estimation (Bergou et al. 2020). To ensure computational robustness, a tryCatch error-handling function was implemented, allowing the script to continue execution even if certain datasets failed to converge (H. Wickham 2014). The fitting process was limited to a maximum of 50 iterations, striking a balance between computational efficiency and accuracy. The evaluation of model performances were conducted using multiple statistical metrics. These metrics are computed for each model variant (Quadratic, Logistic, and Gompertz). This method facilitates the comparative evaluation of all models, quantifying the statistical support each model receives based on its likelihood, which is essential for determining the best-fitting model under the framework of statistical evidence (Royall 1997).

The statistical metrics computed and their explanations:

- **Coefficient of Determination** ($R^2$): This metric quantifies the proportion of variance in the dependent variable that can be explained by the model. A higher $R2^2$ indicates a better model fit, as it suggests that the observed data aligns more closely with the predicted values.

- **Akaike Information Criteria** (AIC): AIC is derived from the Kullback-Leibler (KL) divergence, which measures how much an estimated model deviates from the true data distribution (Akaike 1992). It helps identify the model that minimizes information loss, with lower AIC values indicating a model that best approximates the underlying data-generating process.

- **AICc**:The corrected AIC (AICc) accounts for small sample sizes and is recommended when the ratio of sample size to model parameters ($n/k$) is below 40 (Burnham et al. 2002). Similar to AIC, lower AICc values indicate a model with lower information loss while correcting for bias in small datasets.

- **Bayesian Information Criteria** (BIC): BIC is similar to AIC but applies a stricter penalty for models with additional parameters. This ensures that model complexity is justified by a significantly improved fit to the data, helping to **prevent overfitting (Wagenmakers et al. 2004; Hastie et al. 2001). BIC is particularly useful for comparing models with differing numbers of parameters, such as the **Logistic model (3 parameters) vs. the Gompertz model (4 parameters)**.

## Computing tools

Initially, I employed a Jupyter Notebook to design a workflow that facilitated a preliminary understanding of the dataset's characteristics. All data wrangling and manipulation were performed using Python (v3.12.9) with the pandas library (v2.1.4) for efficient data processing and transformation. I used R (v4.3.3) for model fitting, visualisation, and evaluation with plotting and statistical analysis. Essential packages used included:

- ggplot2 (R): Used data visualisation (Hadley Wickham 2016)

- minpack.lm: Employed for fitting nonlinear curves via the Levenberg-Marquardt nonlinear least-squares algorithm.

- dplyr: Employed in data manipulation and transformation.

- data.table: Provides useful for handling large datasets

- grid: Useful to adjust layouts in ggplot2.

- table: Allows modifications of ggplot2 graphical objects.

- gridExtra: Allowed the multi panel graphics creation.

To compile the results, I used LaTeX and developed a Bash script to automate the execution of these scripts, ensuring both reproducibility and efficiency in data management.

# Results

Model fittings were constrained to data subsets where each model achieved an $R^2$ of at least 0.5, demonstrating a satisfactory fit. The models were evaluated based on several statistical criteria and was considered better with a higher $R^2$, and a lower AIC, AICc, BIC. Each model's "win count" indicates how often it was the best model for a given dataset based on a specific metric Table 2. Among the models evaluated, the natural Quadratic and Logistic Models exhibited moderate performance but were not the most frequently chosen. In contrast, the log Logistic Model performed strongly across various metrics, particularly excelling in AICc. The log Gompertz Model emerged as the most consistently selected model, achieving the highest overall wins across datasets. While the Gompertz Model demonstrated a high $R^2$, it was chosen less frequently when assessed using AIC, AICc, and BIC criteria, suggesting that other models provided a better balance between fit and complexity.

In addition to statistical evaluation, data visualisation was used to further assess model performance (Figure 1). The fitted growth models for Population ID 266 illustrate that the Gompertz (green) and Quadratic (red) models closely follow the observed data, effectively capturing both the exponential growth phase and the stationary phase.

The log Gompertz model (green) also provides a reasonable fit but deviates slightly during the early stages of growth. These results indicate that the Gompertz and Quadratic models may provide the most biologically meaningful fit for this dataset, aligning well with expected population dynamics.

Table 2: Comparison of Model Performance Based on Wins Across Different Metrics. The better model is indicated with a *, as it has the highest amount of "wins".

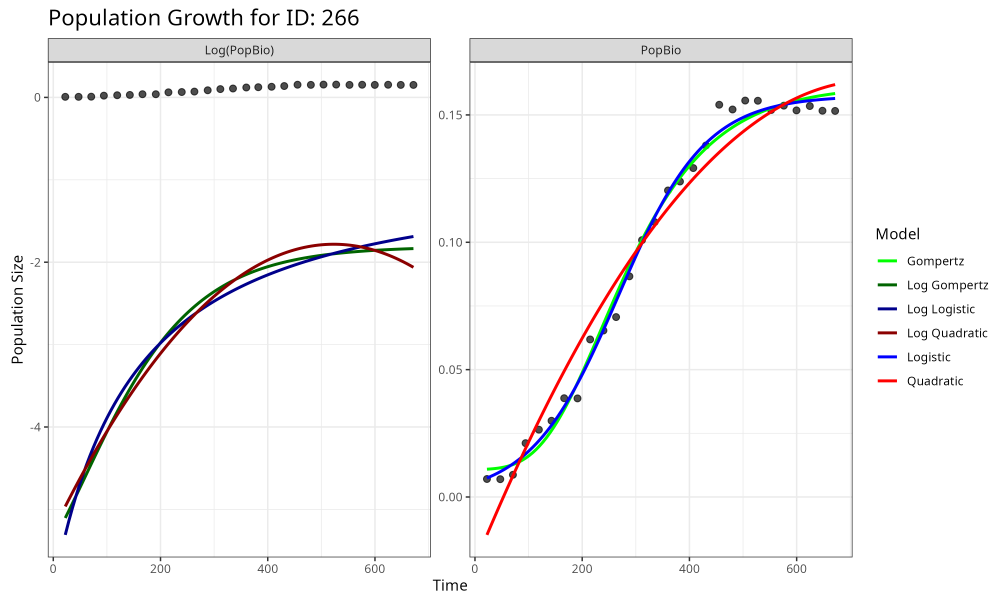| Model | Wins (highest $R^2$) | Wins (lower AIC) | Wins (lower AICc) | Wins (lower BIC) |
|---|---|---|---|---|
| Quadratic | 19 | 30 | 30 | 30 |
| Logistic | 9 | 33 | 33 | 33 |
| Log Quadratic | 8 | 16 | 28 | 16 |
| Log Logistic | 47 | 63 | 114* | 65 |
| Log Gompertz | 124* | 124* | 62 | 122* |
| Gompertz | 80 | 21 | 20 | 21 |

Figure 1: Comparison of Quadratic, Logistic, Gompertz with log-transformed population size (left panel) and un-transformed population size (right panel) for ID 266. The black dots represent observed population growth over time. The Gompertz (green) and Quadratic (red) models effectively represent both the growth phase and the stationary phase. The Log-Gompertz model (green) also provides a good fit but deviates slightly during the initial stages of growth. Dataset from (Bae et al. 2014)

## Discussion

Accurate mathematical models are essential for analysing and predicting bacterial population growth dynamics. Given the diverse growth patterns observed across various species, media, and temperature conditions, the model that provided the best fit in this study can be considered to demonstrate broad applicability for modelling microbial growth (López et al. 2004). Because bacterial growth data can be measured in various ways (OD, CFU, Dry Weight), selecting an appropriate model requires careful consideration of both measurement reliability and statistical properties. Due to heteroscedasticity in CFU and OD measurements, logarithmic transformation is commonly applied (López et al. 2004; Schaffner 1998). I therefore evaluated three models: Quadratic, Logistic, Gompertz models, in both log-transformed and not-transformed Population data.

My findings align with Gibson et al. 1988's approach, as the Gompertz and Logistic models capture microbial growth dynamics with a "reasonable good-ness of fit". In my statistical analysis, the log Gompertz model emerged as the best-fitting model, consistently outperforming others in three out of four statistical criteria, including $R^2$. The log Logistic model had the lowest AICc, which was expected due to it having less parameters (compared to the Gompertz model). These results reinforce the suitability of the Gompertz model for describing microbial growth, particularly when applied to log-transformed data. This is likely because the Gompertz model captures the dynamic nature of growth, where the rate initially increases, reaches a peak, and then declines (Gibson et al. 1988; Jason 1983). However, Figure 1 suggests that statistical fit does not always align

7

with biological interpretability. In the left panel (log-transformed data), model predictions deviate more from observed values, whereas in the right panel (untransformed data), the curves align more closely with actual bacterial growth trends. The Gompertz and Logistic models (green and blue lines) effectively describe both the exponential and stationary phases, reinforcing their relevance in biological contexts.

The choice of model ultimately depends on the objective. For predictive accuracy, the log Gompertz model is optimal, as it explains the greatest variance (highest $R^2$) and performs well across multiple statistical measures. Alternatively, the log Logistic model provides strong predictive performance with lower AICc, indicating a balance between fit and complexity. This may be due to the model's symmetry, as it assumes identical acceleration and deceleration of growth rate, which may not always be biologically accurate (Gibson et al. 1988).

However, for biological realism, the untransformed Gompertz model is preferable, as it explicitly incorporates the lag phase (tlag), a key feature for describing bacterial growth dynamics. If both predictive accuracy and biological interpretability are needed, a phase-specific modelling approach, where different models are applied to distinct growth phases may provide the most comprehensive framework. These results further highlight the need to evaluate growth models not only based on statistical criteria but also how well they align with observed growth patterns. Both statistically and visually, my analysis agrees that the Quadratic model did not directly incorporate biologically meaningful growth parameters and can lead to unrealistic predictions (Gibson et al. 1988). Most of the models I considered fail to account for the bacterial population decline, commonly referred to as the death phase. My study also did not focus on this phase due to the lack of available data, as most datasets I used do not include population measurements during this stage. The exclusion is likely driven by time constraints, financial limitations, and a primary research focus on earlier growth phases, which are particularly relevant in microbiology and antibiotic treatment strategies. The initial growth stage, known as the lag phase, is when microorganisms adjust to their environment and gradually increase their growth potential (Kumakura et al. 2023). A deeper understanding of the mechanisms that enable rapid acclimatisation could facilitate culturing microbes under optimal conditions (Rolfe et al. 2012). Despite its importance to various industries, the lag phase remains the least understood phase of microbial growth due to the scarcity of available data (Rolfe et al. 2012).

Furthermore, the datasets analysed encompassed a range of biological and environmental conditions, including variations in growth media, temperature, and microbial species. As a result, model performance is likely to be influenced by these factors. Even when environmental conditions are controlled, genetic variability within bacterial populations can still lead to substantial differences in growth dynamics (Ackermann 2015). To enhance predictive accuracy, future research should incorporate these biological and environmental variables directly into model frameworks. However, due to the inherent complexity of bacterial growth, no single model is likely to capture all growth phases under all conditions. A more effective approach may involve phase-specific modelling, where

distinct models are tailored to the lag, exponential, stationary, and death phases individually. This strategy could provide a more precise and biologically relevant representation of microbial growth patterns.

Overall, My study underscores the need for a balance between model complexity and practical applicability. While complex models may offer greater accuracy, they must remain interpretable and adaptable to real-world applications. By refining bacterial growth models with phase-specific insights and tailored biological parameters, future research can enhance predictive capabilities, leading to better antimicrobial strategies, biotechnological applications, and ecological modelling of microbial populations.

# References

Ackermann, Martin (2015). "A Functional Perspective on Phenotypic Heterogeneity in Microorganisms". In: *Nature Reviews Microbiology* 13, pp. 497–508.

Akaike, Hirotogu (1992). "Information Theory and an Extension of the Maximum Likelihood Principle". In: *Breakthroughs in Statistics: Foundations and Basic Theory*. Ed. by Samuel Kotz and Norman L. Johnson. New York, NY: Springer New York, pp. 610–624. ISBN: 978-1-4612-0919-5.

Bae, Young-Min et al. (2014). "Growth Characteristics and Biofilm Formation of Various Spoilage Bacteria Isolated from Fresh Produce". In: *Journal of Food Science* 79.10, pp. M2072–M2080.

Bate, Fiona et al. (Aug. 2023). "Emergent Lag Phase in Flux-Regulation Models of Bacterial Growth". English. In: *Bulletin of Mathematical Biology* 85.9. © 2023. The Author(s). ISSN: 1522-9602.

Bergou, E., Y. Diouane, and V. Kungurtsev (2020). *Convergence and Complexity Analysis of a Levenberg-Marquardt Algorithm for Inverse Problems*.

Bernhardt, Joey R., Jennifer M. Sunday, and Mary I. O'Connor (2018). "Metabolic Theory and the Temperature-Size Rule Explain the Temperature Dependence of Population Carrying Capacity". In: *The American Naturalist* 192.6. PMID: 30444656, pp. 687–697.

Brhane, Kewani et al. (Feb. 2024). "Mathematical modeling of cholera dynamics with intrinsic growth considering constant interventions". In: *Scientific Reports* 14.

Buchanan, R.L, R.C Whiting, and W.C Damert (1997). "When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves". In: *Food Microbiology* 14.4, pp. 313–326. ISSN: 0740-0020.

Burnham, K.P. and D.R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Verlag.

Coates, Jessica et al. (Mar. 2018). "Antibiotic-induced population fluctuations and stochastic clearance of bacteria". In: *eLife* 7. Ed. by Aleksandra M Walczak, e32976. ISSN: 2050-084X.

Galarz, Liane Aldrighi, Gustavo Graciano Fonseca, and Carlos Prentice (2016). "Predicting bacterial growth in raw, salted, and cooked chicken breast fillets during storage". In: *Food Science and Technology International* 22.6. PMID: 26683484, pp. 461–474.

Gibson, Angela M., N. Bratchell, and T. A. Roberts (1988). "Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature". In: *International Journal of Food Microbiology* 6.2, pp. 155–178. ISSN: 0168-1605.

Gill, C.O. and K.M. DeLacy (1991). "Growth of Escherichia coli and Salmonella typhimurium on high-pH beef packed under vacuum or carbon dioxide". In: *International Journal of Food Microbiology* 13.1, pp. 21–30. ISSN: 0168-1605.

Gompertz, Benjamin (1825). "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies". In: *Philosophical Transactions of the Royal Society of London* 115, pp. 513–583.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001). *The Elements of Statistical Learning.* Springer Series in Statistics. New York, NY, USA: Springer New York Inc.

Huang, Lihan (2013). "Optimization of a new mathematical model for bacterial growth". In: *Food Control* 32.1, pp. 283–288. ISSN: 0956-7135.

Jason, A.C. (1983). "A deterministic model for monophasic growth of batch cultures of bacteria". In: *Antonie van Leeuwenhoek* 49.6, pp. 513–536.

Jewell, Keith (2012). "Comparison of 1-step and 2-step methods of fitting microbiological models". In: *International Journal of Food Microbiology* 160.2, pp. 145–161. ISSN: 0168-1605.

Koseki, Shigenobu (2016). "Predictive Modeling for Estimation of Bacterial Behavior from Farm to Table". In: *Food Safety* 4.2, pp. 33–44.

Kumakura, Daiki et al. (2023). "Disentangling the growth curve of microbial culture". In: *Journal of Theoretical Biology* 573, p. 111597. ISSN: 0022-5193.

López, S. et al. (2004). "Statistical evaluation of mathematical models for microbial growth". In: *International Journal of Food Microbiology* 96.3, pp. 289–300. ISSN: 0168-1605.

Monod, Jacques (1949). "THE GROWTH OF BACTERIAL CULTURES". In: *Annual Review of Microbiology* 3.Volume 3, 1949, pp. 371–394. ISSN: 1545-3251.

Phillips, J.D. and M.W. Griffiths (1987). "The relation between temperature and growth of bacteria in dairy products". In: *Food Microbiology* 4.2, pp. 173–185. ISSN: 0740-0020.

Pla, María-Leonor et al. (2015). "Comparison of Primary Models to Predict Microbial Growth by the Plate Count and Absorbance Methods". In: *BioMed Research International* 2015.1, p. 365025.

Rolfe, Matthew D. et al. (2012). "Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation". In: *Journal of Bacteriology* 194.3, pp. 686–701.

Roth, Norman G. and Robert B. Wheaton (1962). "CONTINUITY OF PSYCHROPHILIC AND MESOPHILIC GROWTH CHARACTERISTICS IN THE GENUS ¡i¿ARTHROBACTER¡/i¿". In: *Journal of Bacteriology* 83.3, pp. 551–555.

Royall, R.M. (1997). *Statistical Evidence: A Likelihood Paradigm.* Monographs on statistics and applied probability. Chapman & Hall.

Schaffner, D.W. (1998). "Predictive food microbiology Gedanken experiment why do microbial growth data require a transformation?" In: *Food Microbiology* 15.2, pp. 185–189. ISSN: 0740-0020.

Silva, Ana et al. (Oct. 2018). "Modelling the growth of lactic acid bacteria at different temperatures". In: *Brazilian Archives of Biology and Technology* 61.

Singh, Jaskirat Pal et al. (July 2024). "Cholera disease dynamics with vaccination control using delay differential equation". In: *Scientific Reports* 14.

Sivonen, K (1990). "Effects of light, temperature, nitrate, orthophosphate, and bacteria on growth of and hepatotoxin production by Oscillatoria agardhii strains". In: *Applied and Environmental Microbiology* 56.9, pp. 2658–2666.

Stannard, C.J., A.P. Williams, and P.A. Gibbs (1985). "Temperature/growth relationships for psychrotrophic food-spoilage bacteria". In: *Food Microbiology* 2.2, pp. 115–122. ISSN: 0740-0020.

Wagenmakers, Eric-Jan and Simon Farrell (Mar. 2004). "AIC model selection using Akaike weights". In: *Psychonomic bulletin & review* 11, pp. 192–6.

Wickham, H. (2014). *Advanced R*. Chapman & Hall/CRC The R Series. Taylor & Francis. ISBN: 9781466586963.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4.

Winsor, Charles P. (1932). "The Gompertz Curve as a Growth Curve". In: *Proceedings of the National Academy of Sciences* 18.1, pp. 1–8.

Zwietering, M. H. et al. (1994). "Modeling of Bacterial Growth with Shifts in Temperature". In: *Applied and Environmental Microbiology* 60.1, pp. 204–213.