

Does Temperature Influence Model Suitability and Maximum Growth Rates in Microorganisms

Sam Smith

Abstract

1 Introduction

1 Microorganisms make up the most diverse and plentiful group of life on
2 Earth. They are essential for ecosystem function and stability [Shoemaker et al., 2021].
3 In particular microbes play a critical role in processes such as the nitrogen
4 cycle and carbon sequestration that are becoming increasingly significant in
5 light of climate change [Gupta et al., 2016]. Microbes have immense ability
6 to multiply rapidly and exponentially whilst resources are abundant. As nu-
7 trients become limited, Microorganisms must compete thus compromising
8 their ability to reproduce [Nat, 2016]. This results in them typically showing
9 a sigmoidal pattern of growth whereby different stages of their growth curve
10 infer different biology significance. Microbial growth curves typically consist
11 of a lag phase (t_{lag}), exponential growth phase, stationary phase (K), and
12 often a death phase [Zwietering et al., 1990]. The lag phase accounts for the
13 time taken for a population to adjust to its environment before rapid repro-
14 duction [BUC, 1997]. Whereas, the stationary phase or carrying capacity
15 is shown as an asymptote on a bacterial growth curve and indicates often
16 compounding factors such as nutrient availability, predation, and crowding
17 preventing further population growth [Wachenheim et al., 2003].

18

Models can be developed that effectively predict microbial populations at a given time point. They work under the assumption that microbial population patterns are reproducible under the same environmental conditions [Pla et al., 2015]. Modelling microbial population dynamics has positive implications on agriculture and food security as predictions on shelf life and product safety can be made [Zwietering et al., 1990]. In addition, predictive models assist in the decision making of large industrial processes such as fermentation [Garcia et al., 2021]. Models that are derived from existing theory and recorded observations are known as mechanistic; whilst phenomenological models are developed empirically and provide no explanation of patterns [Peleg and Corradini, 2011]. Two examples of commonly used non-linear mechanistic models are the Gompertz model and the Logistic model.

The Logistic equation is popular for describing bacterial population growth. The logistic model that was first developed by Pearl and Reed in 1920 was empirical. However, biological inferences are now commonly derived from its parameters [Wachenheim et al., 2003] [Pearl and Reed, 1920]. Therefore, it can be said that the logistic model estimates the population at any given time point from the initial population (N_0), carrying capacity (K) and maximum growth rate (R_{max}) [Wachenheim et al., 2003]. Whilst models consisting of fewer parameters are often preferred, the logistic equation does lack parameters that captivate other typical stages of microbial growth curves such as a death or lag phase. The Gompertz equation is another sigmoidal model that takes into account the three parameters in the Logistic model but also includes a lag phase. The Gompertz model has been widely used across a range of applications such as predicting plant, fish or even cancerous tumour growth [Tjørve and Tjørve, 2017]. Finally, phenomenological cubic linear models are often used for predicting microbial populations. Garcia et al (2021) found that a cubic model accurately represents all aforementioned stages of bacterial population growth when applied to fermentation bacteria [Garcia et al., 2021]. However, biological inferences cannot be drawn from phenomenological model parameters, highlighting the im-

portance of mechanistic models that provide parameter estimations, as well as predictions.

Statistical analysis whereby different models are selected or ranked in terms of performance have begun to gain traction in ecology and evolution [Johnson and Omland, 2004]. It offers an alternative to traditional hypothesis testing techniques as it confronts multiple 'competing' hypotheses simultaneously. It is crucial to identify the best performing models under various conditions, as the prediction of microbial population dynamics relies on selecting the best-performing model that aligns with the conditions of unsampled populations. This report aims to identify the top performing model (Logistic, Gompertz or Cubic) whilst also examining trends in model performance under different temperatures. It is expected that the Gompertz model may drop in performance relative to the Logistic model with increasing temperatures. This is due to the lag phase being significantly reduced by increasing temperatures [Aba et al., 2021]. Furthermore, I expect no changes in model performance for the Cubic model across temperatures as its parameters are not tied to any biological values that are temperature dependent. Temperature is also known to influence the R_{max} [Ward and Cockson, 1972] [Dey et al., 2020]. Rates of substrate uptake by bacteria reduce with lower temperatures thus negatively impacting a bacteria's ability to grow [Nedwell and Rutter, 1994]. However as both mechanistic models contain an R_{max} parameter this should not vary model performance.

Finally, the model averaging application of model selection techniques will be carried out in this study. Robust parameter estimates for K , N_0 and R_{max} will be calculated using Akaike weights for demonstration and use in potential follow up studies.

- write more on how akaike weights have been used in literature - mention parsimony regarding R^2

84 2 Methods

85 2.1 Data Collection

86 This study is based on an amalgamation of microbial growth curves from 10
87 different research papers. It includes populations of different species grown
88 under varying temperatures (0-37) and 18 different media. Observations in
89 the dataset were deemed from the same curve if they shared a temperature,
90 species and citation. This facilitated the subsetting of the data into 285
91 individual growth curves for model fitting.

92 Model Fitting

93 Only three candidate models were considered in this study as it is ill-advised
94 to include many models that increases the chance of spurious findings [Johnson and Omland, 2004].
95 The Logistic model used in this study is the solution to the differential equation
96 defining the classic logistic population growth equation.

97 2.1.1 Logistic Model

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

98
99 The Gompertz model used in this study is a modified version by Zwieter-
100 ing et al (1990) [Zwietering et al., 1990] where Nmax represents carrying
101 capacity.

102 2.1.2 Gompertz Model

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

103 Finally, the equation below illustrates the cubic equation and its uninter-
104 pretable parameters.

105 2.1.3 Cubic Model

$$\log(N_t) = at + bt^2 + ct^3 + d \quad (3)$$

106

107

108 Fitting the non linear mechanistic models (Logistic and Gompertz) re-
 109 quired reasonable starting values in addition to suitable upper and lower
 110 bounds. Sampling was used to vary the starting values around an appro-
 111 priate mean to increase the likelihood of a non-linear least squares (NLLS)
 112 model fit after 100 attempts. After the 100 attempts for each growth curve,
 113 only the fit with the highest R2 value was outputted alongside its respective
 114 AICc, BIC, and Akaike weight score. All three models were managed to be
 115 fit to 202 out of 285 growth curves. For consistency, when working out the
 116 statistical metrics for the logistic model, the residuals were log transformed
 117 to facilitate comparison between the other models. In addition, log trans-
 118 formed population data is preferred as population is a multiplicative process
 119 therefore residuals naturally increase with time in a linear scale. This will
 120 increase the normality of the model residuals which in turn increases the
 121 predictive power of the models [Freckleton, 2002].

122 2.2 Model Comparison and Weighted Averages

123 Johnson and Omland outlined model selection metrics and described the
 124 differences between them. Firstly, R2 is used that simply suggests the pro-
 125 portion of the variance in the data that is explained by a given model.
 126 However, R squared is not commonly used and is described as 'naive' as it
 127 does not consider model complexity and therefore does not penalise over-
 128 fitted models [Johnson and Omland, 2004]. Secondly, Akaike information
 129 criterion (AIC) is another measure of model performance that counts for
 130 goodness of fit and model complexity. Furthermore, AICc includes bias cor-
 131 rection when the model is applied to small sample sizes. Finally bayesian
 132 information criterion (BIC), similar to AICc, considers fit, complexity and
 133 sample size. However, AIC is often favoured as it is based on Kullback-
 134 Leibler information theory [Johnson and Omland, 2004] - although many
 135 statisticians argue that BIC is preferable as it is less tolerant of overcom-
 136 plex models. When comparing model AIC and BIC values, models with

the smallest AIC and BIC by at least 2, are considered the better performing model [Johnson and Omland, 2004]. Model comparison occurs between models fit onto the same data and an overall model winner is defined as the model that performed best the most number of times. In addition, model performances were compared under all different temperature values to identify trends in model performance.

Omland and Johnson further explain the use of Akaike weights that consider the relative probabilities of a model being the best performer. In this study, for a model to be selected as the best performer an Akaike weighted score of over 0.9 is necessary [Dash et al., 2023]. This is an arbitrary threshold that suggests that a model can only be deemed the best if there is more than 90 percent chance that that is the case. These probabilities can consequently be used for robust parameter estimations. In this study, Akaike weighted averages were made for comparison between all three models. However, for parameter estimations Akaike weights were recalculated for just Gompertz and Logistic as the Cubic model does not contain interpretable parameters associated with stages of bacterial growth.

2.3 Computational Tools

3 Results

All 202 growth curves with fitted models were plotted with both log transformed and linear population values. For example subset 225 is a good example for illustrating the differences between plot types:

In both plots it is clear that the Gompertz and Cubic models were best fit to the data. This would suggest their R^2 values would be stronger however, as the logistic model contains fewer parameters, AIC and BIC values may tell a different story. Furthermore, both plots illustrate the Logistic models struggle to fit curves with a time lag phase. The Gompertz and Cubic manage to capture the lag phase, growth phase and stationary phase,

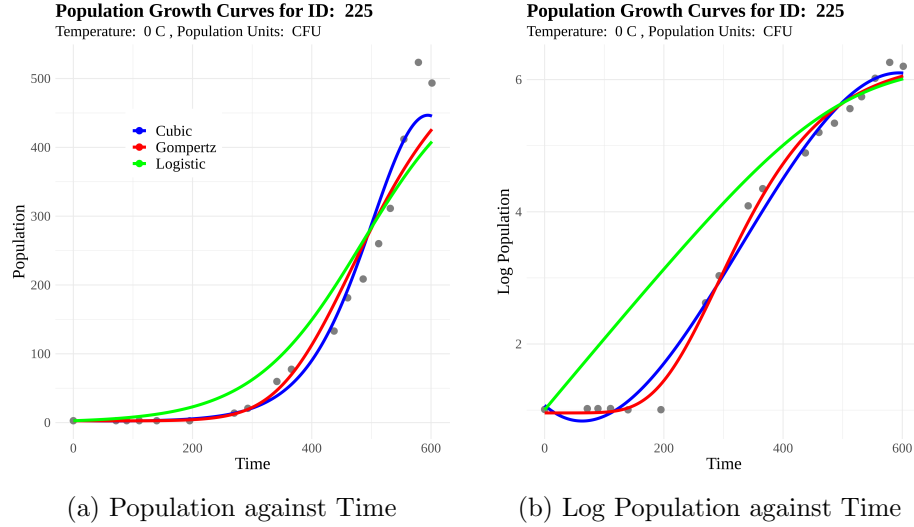


Figure 1: Graphs showing the relationship between log and linear microbial populations against time for the subset 225.

with the cubic even predicting a death phase shown by a slight decline in predicted population.

169

The number of times model performed best for each criteria is given in this table below:

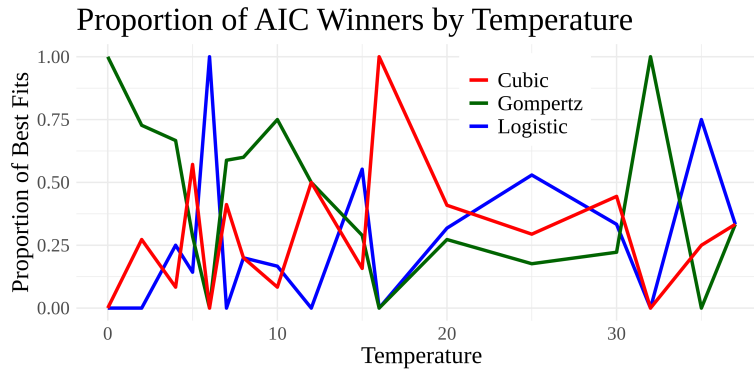
Model	AICc	BIC	Rsqr	AkaikeW
Logistic	56	28	26	47
Cubic	52	73	84	40
Gompertz	72	76	92	59

The table shows that Gompertz was the best performing model the most number of times for each criteria. The Gompertz model's quality of fit clearly outweighed its penalty for having an extra parameter than the Logistic model for AICc and BIC. Furthermore, the Cubic and Gompertz models clearly outperformed the Logistic according to Rsqr, aligning with expectations due to their greater number of parameters. Despite this, the number of Akaike weights best fits is fairly similar across all three models. The

180 Gompertz model had a probability of being the best model of more than 0.9
 181 only 59 times out of the 202 growth curves. It is important to note that the
 182 columns may not add up to 202 due to 'no clear winner' outcomes.

183

184 The proportions of best fits for each model were calculated and plotted
 185 for each unique temperature value in the dataset.



186

187 Observing the plot there is no clear trend between performance for any model
 188 and temperature. Whilst there are temperatures which certain models per-
 189 form best at, for example cubic at 16 degrees, drawing definitive trends is
 190 challenging due to small sample sizes for each temperature.

191 4 Discussion

192 The trade off between parsimony and complexity is an important dilemma
 193 to address in this case. Parsimonious theories advocate for simplicity of
 194 explanation [Coelho et al., 2019] thus the most straightforward model that
 195 adequately explains the data is best. Simplicity is often favoured as it is more
 196 interpretable and overfitting is less likely. In contrast, complex models with
 197 more parameters than its competitors (such as Gompertz) explain intricacies
 198 in the data, thus providing a better fit. Therefore, metrics that quantify
 199 the balance of parsimony and complexity, AIC and BIC, were used in this
 200 study to identify the best performing models.

201

202 In this study we identified the Gompertz model to be the best predictor
 203 of microbial population growth - implying that the additional parameter
 204 compared to the Logistic model provide sufficient explanation of the data
 205 to justify its complexity. Moreover, there was no evidence from figure X
 206 to suggest that the efficacy all models changed with temperature. There-
 207 fore, this report encourages the use of the Gompertz model for predicting
 208 microbial populations across a range of temperatures. The results do not
 209 align with the expectation that the Logistic model would outperform the
 210 Gompertz model at higher temperatures. Reasons for this may be due to
 211 the limitations of this study such as the other confounding factors at play.
 212 Firstly, the length and extent of different stages of microbial growth will de-
 213 pend on its environmental conditions and the environmental conditions that
 214 it is best adapted for [Dey et al., 2020]. For example, lag phase increases for
 215 microbes placed in novel environments to allow for time taken to adapt to
 216 new conditions [Rolfe et al., 2012]. This favours the Gompertz model per-
 217 formance where a lag parameter is present. Secondly, growth mediums are
 218 known to affect metabolic rates in bacteria [Kim and Kim, 2017]. Kim and
 219 Kim (2017) found that microbes grown in nutrient rich media had increased
 220 K and Rmax and decreased lags than those grown in nutrient poor me-
 221 dia. This indicates that the Logistic equation is more suitable for nutrient
 222 rich media. Thirdly, different species of microorganisms exhibit different
 223 patterns of growth. For example, *Tetraselmis tetrahele*, a species of phy-
 224 toplankton included in the study is likely to have a lag phase synonymous
 225 with most algal species [Collos, 1986]. This increases the feasibility of the
 226 Gompertz and Cubic models. To address the limitations of this study, more
 227 specified experiments that record enough data for each unique set of condi-
 228 tions would provide a more nuanced understanding of model performance.

229
 230 The limited sized data set also reduces the significance of the findings
 231 of the study. Whilst the Gompertz model performed best for all criteria,
 232 no statistical analyses were conducted to address whether or not Gompertz
 233 would likely be the best performing model in another set of growth curves.
 234 Furthermore, the Akaike weight scores illustrate no overwhelming winner

235 and therefore suggest model selection should be more specific to the con-
 236 ditions that a microbe is grown under. For example, when predicting the
 237 parameters or population growth of bacteria grown under the same condi-
 238 tions it is adapted to then the Logistic model should be selected for reasons
 239 previously mentioned. On the other hand, when handling microbes grown
 240 under nutrient poor environment then a model with a lag phase param-
 241 eter is more suitable. This study did not contain a plentiful enough data
 242 set to select the best models for each unique set of environmental condi-
 243 tions. However, to combat this, the Akaike weighted parameter estimations
 244 could be used to predict microbial population growth - particularly when
 245 there is no clearly best performing model. Model averaging using Akaike
 246 weights is mostly useful when multiple models exhibit roughly equal AIC
 247 values [Johnson and Omland, 2004], so equally performing models are both
 248 considered when calculating parameters. This study generated Akaike esti-
 249 mated parameter values regardless of whether or not there was a clear best
 250 performer ($Aw \geq 0.9$). However, a potential improvement to the estima-
 251 tions, as suggested by Johnson and Omland (2004), would be to only use
 252 model averaging when the best performing models Akaike weight is less than
 253 0.9. And otherwise just use given parameter of the best performing model.
 254 This is to prevent the influence of poor performing models on parameter
 255 estimates.

256 **5 Conclusion**

257 In conclusion the results of this study provides explanation for why the Gom-
 258 pertz model should be used to gain a general idea of microbial population
 259 growth. However, it is likely that selecting models that perform best for
 260 particular conditions, or using akaike weights is more reliable method for
 261 predicting growth curves and parameters.

References

- [BUC, 1997] (1997). 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.
- [Nat, 2016] (2016). *Nature Reviews Microbiology*, 14(9):543–543.
- [Aba et al., 2021] Aba, R. P. M., Gelido, E. M. L., Yatco, K. M. R. S., and Gabriel, A. A. (2021). Microbial shelf life of coconut water subjected to various inoculation levels of listeria monocytogenes and storage conditions. *International Journal of Food Microbiology*, 344:109108.
- [Coelho et al., 2019] Coelho, M. T. P., Diniz-Filho, J. A., and Rangel, T. F. (2019). A parsimonious view of the parsimony principle in ecology and evolution. *Ecography*, 42(5):968–976.
- [Collos, 1986] Collos, Y. (1986). Time-lag algal growth dynamics: biological constraints on primary production in aquatic environments. *Marine Ecology Progress Series*, 33(2):193–206.
- [Dash et al., 2023] Dash, G., Sen, S., Pradhan, R. K., Ghosh, S., and Jose, J. (2023). Application of information theory-based decision support system for high precision modeling of the length-weight relationship (lwr) for five marine shrimps from the northwestern bay of bengal. *Regional Studies in Marine Science*, 66:103140.
- [Dey et al., 2020] Dey, A., Bokka, V., and Sen, S. (2020). Dependence of bacterial growth rate on dynamic temperature changes. *IET Systems Biology*, 14(2):68–74.
- [Freckleton, 2002] Freckleton, R. P. (2002). On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *Journal of Animal Ecology*, 71(3):542–545.
- [Garcia et al., 2021] Garcia, B., Rodriguez, E., Salazar, Y., Valle, P., Flores-Gallegos, A., Rutiaga-Quñones, O., and Rodriguez-Herrera, R. (2021).

290 Primary model for biomass growth prediction in batch fermentation. *Sym-*
291 *metry*, 13(8):1468.

292 [Gupta et al., 2016] Gupta, A., Gupta, R., and Singh, R. L. (2016). *Mi-*
293 *crobes and Environment*, page 43–84. Springer Singapore.

294 [Johnson and Omland, 2004] Johnson, J. B. and Omland, K. S. (2004).
295 Model selection in ecology and evolution. *Trends in Ecology Evolution*,
296 19(2):101–108.

297 [Kim and Kim, 2017] Kim, J. and Kim, K. H. (2017). Effects of minimal
298 media vs. complex media on the metabolite profiles of escherichia coli and
299 saccharomyces cerevisiae. *Process Biochemistry*, 57:64–71.

300 [Nedwell and Rutter, 1994] Nedwell, D. B. and Rutter, M. (1994). Influence
301 of temperature on growth rate and competition between two psychrotol-
302 erant antarctic bacteria: low temperature diminishes affinity for substrate
303 uptake. *Applied and Environmental Microbiology*, 60(6):1984–1992.

304 [Pearl and Reed, 1920] Pearl, R. and Reed, L. J. (1920). On the rate of
305 growth of the population of the united states since 1790 and its mathe-
306 matical representation. *Proceedings of the National Academy of Sciences*,
307 6(6):275–288.

308 [Peleg and Corradini, 2011] Peleg, M. and Corradini, M. G. (2011). Mi-
309 crobial growth curves: What the models tell us and what they cannot.
310 *Critical Reviews in Food Science and Nutrition*, 51(10):917–945.

311 [Pla et al., 2015] Pla, M.-L., Oltra, S., Esteban, M.-D., Andreu, S., and
312 Palop, A. (2015). Comparison of primary models to predict microbial
313 growth by the plate count and absorbance methods. *BioMed Research*
314 *International*, 2015:1–14.

315 [Rolfe et al., 2012] Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thomp-
316 son, A., Cameron, A. D. S., Alston, M., Stringer, M. F., Betts, R. P.,
317 Baranyi, J., Peck, M. W., and Hinton, J. C. D. (2012). Lag phase is
318 a distinct growth phase that prepares bacteria for exponential growth

319 and involves transient metal accumulation. *Journal of Bacteriology*,
320 194(3):686–701.

321 [Shoemaker et al., 2021] Shoemaker, W. R., Jones, S. E., Muscarella, M. E.,
322 Behringer, M. G., Lehmkuhl, B. K., and Lennon, J. T. (2021). Microbial
323 population dynamics and evolutionary outcomes under extreme energy
324 limitation. *Proceedings of the National Academy of Sciences*, 118(33).

325 [Tjørve and Tjørve, 2017] Tjørve, K. M. C. and Tjørve, E. (2017). The
326 use of gompertz models in growth analyses, and new gompertz-model
327 approach: An addition to the unified-richards family. *PLOS ONE*,
328 12(6):e0178691.

329 [Wachenheim et al., 2003] Wachenheim, D. E., Patterson, J. A., and
330 Ladisch, M. R. (2003). Analysis of the logistic function model: derivation
331 and applications specific to batch cultured microorganisms. *Bioresource*
332 *Technology*, 86(2):157–164.

333 [Ward and Cockson, 1972] Ward, J. and Cockson, A. (1972). Studies on a
334 thermophilic bacillus: Its isolation, properties, and temperature coeffi-
335 cient of growth. *Journal of Bacteriology*, 112(3):1040–1042.

336 [Zwietering et al., 1990] Zwietering, M. H., Jongenburger, I., Rombouts,
337 F. M., and van ’t Riet, K. (1990). Modeling of the bacterial growth
338 curve. *Applied and Environmental Microbiology*, 56(6):1875–1881.