

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 [?]. In par-
3 ticular microbes play a critical role in processes such as the nitrogen cycle
4 and carbon sequestration that are becoming increasingly significant in light
5 of climate change [?]. Microbes have immense ability to multiply rapidly and
6 exponentially whilst resources are abundant. As nutrients become limited,
7 microorganisms must compete thus compromising their ability to reproduce
8 [?]. This results in them typically showing a sigmoidal pattern of growth
9 whereby different stages of their growth curve infer different biology sig-
10 nificance. Microbial growth curves typically consist of a lag phase (t_{lag}),
11 exponential growth phase, stationary phase (K), and often a death phase
12 [?]. The lag phase accounts for the time taken for a population to adjust
13 to its environment before rapid reproduction [?]. Whereas, the stationary
14 phase or carrying capacity is shown as an asymptote on a bacterial growth
15 curve and indicates often compounding factors such as nutrient availability,
16 predation, and crowding preventing further population growth [?]. The max-
17 imum growth rate (R_{max}) is also typically annotated on microbial growth
18 curves, illustrated by the steepest gradient.

19

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 [?]. Modelling microbial population dynamics has positive implications on agriculture and food security as predictions on shelf life and product safety can be made [?]. In addition, predictive models assist in the decision making of large industrial processes such as fermentation [?]. 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 [?]. Two examples of commonly used non-linear mechanistic models are the 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 [?] [?]. Therefore, it can be said that the logistic model estimates the population at any given time point from the initial population (N_0), K and R_{max} values [?]. 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 tlag. The Gompertz model has been widely used across a range of applications such as predicting plant, fish or even cancerous tumour growth [?]. 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 [?]. However, biological inferences cannot be drawn from phenomenological model parameters, highlighting the importance of mechanistic models that provide parameter estimations, as well as predictions.

Statistical analysis whereby different models are selected or ranked in

53 terms of performance have begun to gain traction in ecology and evolution
 54 [?]. It offers an alternative to traditional hypothesis testing techniques as
 55 it confronts multiple 'competing' hypotheses simultaneously. It is crucial to
 56 identify the best performing models under various conditions, as the predic-
 57 tion of microbial population dynamics relies on selecting the best-performing
 58 model that aligns with the conditions of unsampled populations. This re-
 59 port aims to identify trends in model performance between Cubic, Gompertz
 60 and Logistic, under different temperatures. It is expected that the Gompertz
 61 model may drop in performance relative to the Logistic model with increas-
 62 ing temperatures. This is due to the lag phase being significantly reduced
 63 by increasing temperatures [?]. Furthermore, I expect no changes in model
 64 performance for the Cubic model across temperatures as its parameters are
 65 not tied to any biological values that are temperature dependent. Temper-
 66 ature is also known to influence the R_{max} [?] [?]. Rates of substrate uptake
 67 by bacteria reduce with lower temperatures thus negatively impacting a bac-
 68 teria's ability to grow [?]. However as both mechanistic models contain an
 69 R_{max} parameter this should not vary model performance.

70

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

75 2 Methods

76 2.1 Data Collection

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

83 **Model Fitting**

84 Only three candidate models were considered in this study as it is ill-advised
85 to include many models that increases the chance of spurious findings [?].
86 The Logistic model used in this study is the solution to the differential
87 equation defining the classic logistic population growth equation.

88 **2.1.1 Logistic Model**

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

89 The Gompertz model used in this study is a modified version by Zwietering
90 et al (1990) [?] where Nmax represents carrying capacity.

91 **2.1.2 Gompertz Model**

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

92 Finally, the equation below illustrates the cubic equation and its uninter-
93 pretable parameters.

94 **2.1.3 Cubic Model**

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

95 Fitting the non linear mechanistic models (Logistic and Gompertz) re-
96 quired reasonable starting values in addition to suitable upper and lower
97 bounds. Sampling was used to vary the starting values around an appro-
98 priate mean to increase the likelihood of a non-linear least squares (NLLS)
99 model fit after 100 attempts. After the 100 attempts for each growth curve,
100 only the fit with the highest R2 value was outputted alongside its respective
101 AICc, BIC, and Akaike weight score. All three models were managed to be
102 fit to 202 out of 285 growth curves. For consistency, when working out the
103 statistical metrics for the logistic model, the residuals were log transformed
104 to facilitate comparison between the other models. In addition, log trans-
105 formed population data is preferred as population is a multiplicative process

106 therefore residuals naturally increase with time in a linear scale. This will
107 increase the uniformity of the model residuals which leads to consistent
108 predictive power across the range of time values [?].

109 **2.2 Model Comparison and Weighted Averages**

110 Johnson and Omland outlined model selection metrics and described the
111 differences between them. Firstly, R^2 is used that simply suggests the pro-
112 portion of the variance in the data that is explained by a given model. R
113 squared is not commonly used and is described as 'naive' as it does not con-
114 sider model complexity and therefore does not penalise overfitted models [?].
115 Secondly, Akaike information criterion (AIC) is another measure of model
116 performance that counts for goodness of fit and model complexity. Fur-
117 thermore, AICc includes bias correction when the model is applied to small
118 sample sizes. Finally bayesian information criterion (BIC), similar to AICc,
119 considers fit, complexity and sample size. However, AIC is often favoured as
120 it is based on Kullback-Leibler information theory [?] - though some statis-
121 ticians still argue that BIC is preferable as it is less tolerant of overcomplex
122 models. When comparing model AIC and BIC values, models with the
123 smallest AIC and BIC by at least 2, are considered the better performing
124 model [?]. Model comparison occurs between models fit onto the same data
125 and an overall model winner is defined as the model that performed best the
126 most number of times. In addition, model performances were compared un-
127 der all different temperature values to identify trends in model performance.

128

129 Omland and Johnson further explain the use of a more interpretable
130 metric, Akaike weights that consider the relative probabilities of a model
131 being the best performer. In this study, for a model to be selected as the
132 best performer an Akaike weighted score of over 0.9 is necessary [?]. This
133 is an arbitrary threshold suggesting a model can only be deemed the best
134 if there is more than 90 percent chance that that is the case. These proba-
135 bilities can consequently be used for robust parameter estimations. In this
136 study, Akaike weighted averages were made for comparison between all three

137 models. However, for parameter estimations Akaike weights were recalcu-
138 lated for just Gompertz and Logistic as the Cubic model does not contain
139 interpretable parameters associated with stages of bacterial growth.

140 **2.3 Computational Tools**

141 Firstly the pandas package in python 3.10.12 [?] was used for initial data ma-
142 nipulation and exportation as it is a computationally efficient and popular
143 tool. R version 4.1.2 [?] was consequently used for model fitting, analy-
144 sis, and plotting. For NLLS model fitting, minpack.lm package was used
145 to access the nlsLM function that uses the Levenberg-Marquardt algorithm,
146 known to be more robust than the Gauss-Newton algorithm. The purrr
147 package, which incorporates the possibly function, was employed to con-
148 tinue the iterative process in the model fitting and plotting functions in the
149 presence of errors. The dplyr package was used for appending model win-
150 ners to respective IDs before tables and graphs could be plotted. Finally,
151 ggplot2 was used for producing all graphs included in the report.

152 **3 Results**

153 All 202 growth curves with fitted models were plotted with both log trans-
154 formed and linear population values. Subset 225 is a good example for
155 illustrating the differences between plot types:

156 In both plots it is clear that the Gompertz and Cubic models were best
157 fit to the data. This would suggest their R2 values would be stronger how-
158 ever, as the logistic model contains fewer parameters, AIC and BIC values
159 may tell a different story. Furthermore, both plots illustrate the Logistic
160 models struggle to fit curves with a time lag phase. The Gompertz and Cu-
161 bic manage to captivate the lag phase, growth phase and stationary phase,
162 with the cubic even predicting a death phase shown by a slight decline in
163 predicted population.

164

165 The number of times each model performed best for a given criteria is

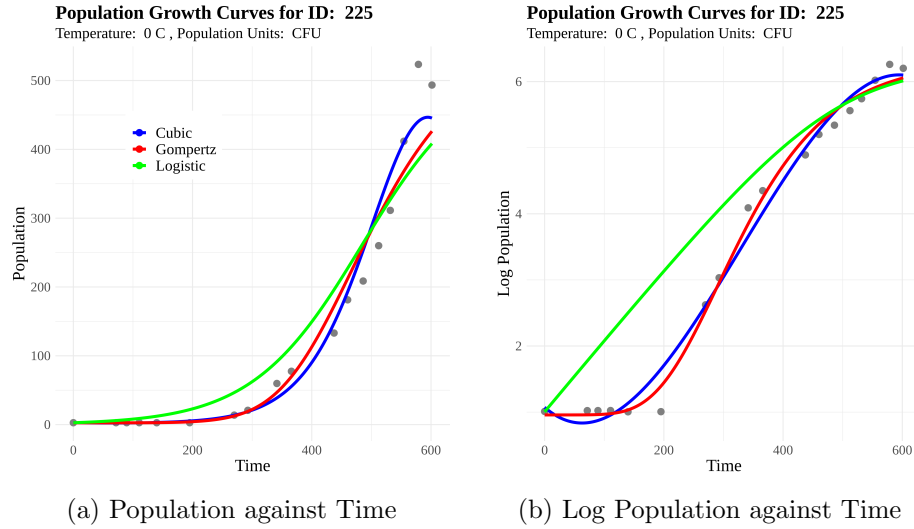


Figure 1: Graphs showing the relationship between log and linear microbial populations against time for subset 225.

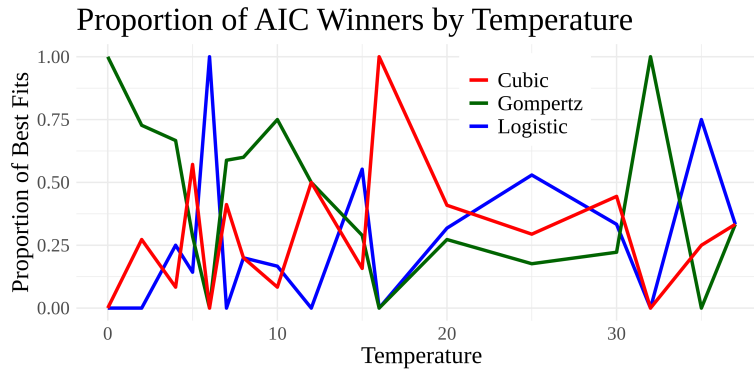
shown in this table below:

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

179

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



182

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

187 4 Discussion

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

197

198 In this study we identified the Gompertz model to be the best predictor
 199 of microbial population growth - implying that the additional parameter
 200 compared to the Logistic model provides sufficient explanation of the data

201 to justify it's complexity. Moreover, there was no evidence from figure X to
202 suggest that the efficacy all models changed with temperature. Therefore,
203 this report encourages the use of the Gompertz model for predicting micro-
204 bial populations across a range of temperatures.

205

206 The results do not align with the expectation that the Logistic model
207 would outperform the Gompertz model at higher temperatures. Reasons
208 for this may be due to the limitations of this study such as the other con-
209 founding factors at play. Firstly, the length and extent of different stages
210 of microbial growth will depend on the environmental conditions relative to
211 the conditions that it is best adapted for [?]. For example, tlag increases for
212 microbes placed in novel environments to allow for time taken to adapt to
213 new conditions [?]. This favours the Gompertz model where a tlag param-
214 eter is present. Secondly, growth mediums are known to affect metabolic
215 rates in bacteria [?]. Kim and Kim (2017) found that microbes grown in nu-
216 trient rich media had increased K and Rmax and decreased tlags than those
217 grown in nutrient poor media. This indicates that the Logistic equation is
218 more suitable for nutrient rich media. This increases the feasibility of the
219 Gompertz and Cubic models. To address the limitations of this study, more
220 specified experiments that record enough data for each unique set of condi-
221 tions would provide a more nuanced understanding of model performance.

222

223 The limited data set also reduces the significance of the findings of the
224 study. Whilst the Gompertz model performed best for all criteria, no sta-
225 tistical analyses were conducted to address whether or not Gompertz would
226 likely be the best performing model in another set of growth curves. Fur-
227 thermore, the Akaike weight scores illustrate no overwhelming winner and
228 therefore suggest model selection should be more specific to the conditions
229 that a microbe is grown under. For example, when predicting the param-
230 eters or population growth of bacteria grown under the same conditions
231 it is adapted to, then the Logistic model should be selected for reasons
232 previously mentioned. On the other hand, when handling microbes grown
233 under nutrient poor environment then a model with a lag phase parameter

234 is more suitable. This study did not contain a plentiful enough data set
 235 to select the best models for each unique set of environmental conditions.
 236 However, to combat this, the Akaike weighted parameter estimations could
 237 be used to predict microbial population growth - particularly when there is
 238 no clearly best performing model. Model averaging using Akaike weights is
 239 mostly useful when multiple models exhibit roughly equal AIC values [?],
 240 so equally performing models are both considered when calculating param-
 241 eters. This study generated Akaike estimated parameter values regardless
 242 of whether or not there was a clear best performer ($Aw \geq 0.9$). However,
 243 a potential improvement to the estimations, as suggested by Johnson and
 244 Omland (2004), would be to only use model averaging when the best per-
 245 forming models Akaike weight is less than 0.9. And otherwise just use given
 246 parameter of the best performing model. This is to prevent the influence of
 247 poor performing models on parameter estimates.

248 **5 Conclusion**

249 In conclusion the results of this study provides explanation for why the
 250 Gompertz model should be used to gain a general idea of microbial popu-
 251 lation growth. However, it is likely that selecting models that perform best
 252 for particular conditions, or using akaike weights is more reliable method
 253 for predicting growth curves and parameters. The study also generated
 254 robust parameter estimates for Rmax and K (found in the supplementary
 255 information) that may be used in future studies.

256 **References**

- 257 [BUC, 1997] (1997). When is simple good enough: a comparison of the gom-
 258 pertz, baranyi, and three-phase linear models for fitting bacterial growth
 259 curves. *Food Microbiology*, 14(4):313–326.
 260 [Nat, 2016] (2016). *Nature Reviews Microbiology*, 14(9):543–543.

- 261 [Aba et al., 2021] Aba, R. P. M., Gelido, E. M. L., Yatco, K. M. R. S., and
 262 Gabriel, A. A. (2021). Microbial shelf life of coconut water subjected
 263 to various inoculation levels of listeria monocytogenes and storage condi-
 264 tions. *International Journal of Food Microbiology*, 344:109108.
- 265 [Coelho et al., 2019] Coelho, M. T. P., Diniz-Filho, J. A., and Rangel, T. F.
 266 (2019). A parsimonious view of the parsimony principle in ecology and
 267 evolution. *Ecography*, 42(5):968–976.
- 268 [Dash et al., 2023] Dash, G., Sen, S., Pradhan, R. K., Ghosh, S., and Jose,
 269 J. (2023). Application of information theory-based decision support sys-
 270 tem for high precision modeling of the length-weight relationship (lwr)
 271 for five marine shrimps from the northwestern bay of bengal. *Regional*
 272 *Studies in Marine Science*, 66:103140.
- 273 [Dey et al., 2020] Dey, A., Bokka, V., and Sen, S. (2020). Dependence of
 274 bacterial growth rate on dynamic temperature changes. *IET Systems*
 275 *Biology*, 14(2):68–74.
- 276 [Freckleton, 2002] Freckleton, R. P. (2002). On the misuse of residuals in
 277 ecology: regression of residuals vs. multiple regression. *Journal of Animal*
 278 *Ecology*, 71(3):542–545.
- 279 [Garcia et al., 2021] Garcia, B., Rodriguez, E., Salazar, Y., Valle, P., Flores-
 280 Gallegos, A., Rutiaga-Quñones, O., and Rodriguez-Herrera, R. (2021).
 281 Primary model for biomass growth prediction in batch fermentation. *Sym-*
 282 *metry*, 13(8):1468.
- 283 [Gupta et al., 2016] Gupta, A., Gupta, R., and Singh, R. L. (2016). *Mi-*
 284 *crobes and Environment*, page 43–84. Springer Singapore.
- 285 [Johnson and Omland, 2004] Johnson, J. B. and Omland, K. S. (2004).
 286 Model selection in ecology and evolution. *Trends in Ecology Evolution*,
 287 19(2):101–108.

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

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

295 [Pearl and Reed, 1920] Pearl, R. and Reed, L. J. (1920). On the rate of
296 growth of the population of the united states since 1790 and its mathe-
297 matical representation. *Proceedings of the National Academy of Sciences*,
298 6(6):275–288.

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

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

306 [R Core Team, 2021] R Core Team (2021). *R: A Language and Environ-*
307 *ment for Statistical Computing*. R Foundation for Statistical Computing,
308 Vienna, Austria.

309 [Rolfe et al., 2012] Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thomp-
310 son, A., Cameron, A. D. S., Alston, M., Stringer, M. F., Betts, R. P.,
311 Baranyi, J., Peck, M. W., and Hinton, J. C. D. (2012). Lag phase is
312 a distinct growth phase that prepares bacteria for exponential growth
313 and involves transient metal accumulation. *Journal of Bacteriology*,
314 194(3):686–701.

315 [Shoemaker et al., 2021] Shoemaker, W. R., Jones, S. E., Muscarella, M. E.,
316 Behringer, M. G., Lehmkuhl, B. K., and Lennon, J. T. (2021). Microbial

317 population dynamics and evolutionary outcomes under extreme energy
318 limitation. *Proceedings of the National Academy of Sciences*, 118(33).

319 [Tjørve and Tjørve, 2017] Tjørve, K. M. C. and Tjørve, E. (2017). The
320 use of gompertz models in growth analyses, and new gompertz-model
321 approach: An addition to the unified-richards family. *PLOS ONE*,
322 12(6):e0178691.

323 [Van Rossum and Drake, 2009] Van Rossum, G. and Drake, F. L. (2009).
324 *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

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

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

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