

2 How well do different mathematical models  
3 based upon population growth  
4 (mechanistic) theory vs. phenomenological  
5 ones, fit to microbial growth data?

6 Author: Kayleigh Greenwood

7 Internal Supervisor: Dr James Rosindell, Imperial College London (j.rosindell@imperial.ac.uk)

8 External Supervisor: Dr Joss Wright, University of Oxford (joss.wright@oii.ox.ac.uk)

9 04/04/2022

## 10 Introduction

11 Microbial growth rates are important to understand in the context of food safety, as there are signifi-  
12 cant financial and health burdens of foodborne illnesses [Daniel et al., 2020], [Organization et al., 2015].  
13 It is important to be able to predict microbial growth as accurately as possible in order to reduce these  
14 impacts via precise estimates of shelf life and other characteristics of growth [McMeekin and Ross, 1996].  
15 The standard method for growth prediction used to be null hypothesis testing, however an increas-  
16 ingly popular alternative is mathematical modelling [Foegeding, 1997]. Growth models typically con-  
17 tain four phases; lag, exponential, stationary and death. The death phase is typically excluded in the  
18 context of food microbiology, as because food is almost certainly spoiled or unfit by the time this phase  
19 begins, it is irrelevant [Ross and McMeekin, 2003], and therefore will not be considered throughout  
20 this paper. This approach consists of fitting various models to data, each representing a separate  
21 hypothesis, and using model selection to determine which model is the best. Model fitting is benef-  
22 cial as it allows multiple hypothesis to be assessed and compared at once, as opposed to only being  
23 able to accept or reject a null hypothesis. Levins (1966) highlighted that the ideal model will never be  
24 feasible, forming the basis of the complexity of modelling, and why there is no one uniform population  
25 model which spans population biology. Literature on microbial growth rates contains various mod-  
26 els with countless combinations of assumptions and techniques. Commonly used empirical models  
27 prioritise realism and precision but sacrifice generality. On the other hand, popular general models  
28 which prioritise generality and make a trade off with precision.

29 Both empirical and mechanistic models have been used throughout the literature for modelling  
30 microbial growth. Initially, empirical models were used which were derived from models outside of the  
31 microbial growth sector, such as the Gompertz and Logistic models. However, mechanistic models  
32 have since been developed, such as the Baryani model [Grijnspeerdt and Vanrolleghem, 1999]. The  
33 benefits of recently developed mechanistic models are that the model's parameters have a theoretical  
34 basis, however, there will always be doubt around whether the underlying mechanisms that the results  
35 rely on are accurate [López et al., 2004]. This gives empirical models an advantage in that they don't  
36 have any risk associated with mechanistic assumptions, and although they have no theoretical basis,  
37 it is argued that explanation of a relationship/behaviour isn't necessary in order to predict it, and  
38 sometimes getting a prediction as precise as possible is the most important thing.

39 The contrast described above is why neither empirical nor mechanistic models have prevailed  
40 over the other. In fitting both types of models to many samples within this dataset, I plan on using  
41 various model comparison/selection techniques to identify the strengths and weaknesses of the two  
42 modelling approaches in the context of microbial growth rates. By identifying the unique statistical  
43 aspects of each model I have aimed to discover if empirical or mechanistic models are best suited to  
44 microbial growth hypotheses.

## 45 Methods

### 46 Data

47 I used an existing database to model microbial growth rate data in this study. The logistic growth  
48 data allowed me to draw conclusions about the studied models' compatibility with microbial growth  
49 rates by modelling bacterial abundance as the response variable, against time as the explanatory  
50 variable. This explanatory/response dynamic in microbial biology allows conclusions and theories to

51 focus on the extent to which time affects microbial growth, and how big of an effect this is. The data  
52 contains measures of bacterial abundance over time, across various combinations of other variables.  
53 The dataset comprises 4387 observations across 10 variables, however any entries with negative  
54 time or abundance values were excluded, leaving 4294 observations for further analysis. Sample  
55 groups were sub-divided according to species, temperature, medium and experiment, resulting in  
56 277 subsets.

## 57 **Models**

58 To test the suitability of empirical, linear models for microbial growth rate studies, I used a cubic  
59 polynomial. The exponential nature of growth means that the log is normally plotted in order to  
60 normalize variance, causing the data to behave in a sigmoidal fashion [Zwietering et al., 1990]. A  
61 cubic polynomial has two peaks, and therefore three separate areas, which is the same amount of  
62 distinct areas as a growth curve has when studied in microbial growth, suggesting fitting this model  
63 would work well as the shape of a cubic curve resembles the shape of a sigmoidal behaviour of  
64 the growth rates. For the representative mechanistic model, I chose to fit the Logistic model. The  
65 Logistic model is a non-linear model, which I chose to plot because of its competitive advantage in  
66 the estimation of when the death phase has been reached (the maximum value/carrying capacity).  
67 Although the death phase itself is not explicitly relevant to microbial growth, a parameter estimate  
68 of carrying capacity as accurate as what this model produces has importance in further analysis  
69 abilities. Being able to predict carrying capacity as well as the Logistic Model does means that  
70 estimates of other values and parameters beyond this analysis are improved.

## 71 **Model fitting**

72 I fitted the linear model with ordinary least squares, and used NLLS for the non-linear model. I used  
73 the nlsLM function in R to fit the non-linear model, within the minpack.lm package.

## 74 **Model selection**

75 For each plotted subset which fit on it both the linear and non-linear model I used AIC values as a  
76 method of comparing the fit of the two models. Given that a lower AIC is better, I will use this and a  
77 general rule of an AIC value difference of 2 to signify significant difference, on each plotted subset  
78 as an indicator of model fit. Additionally, I used residual sums of squares to determine which model  
79 explained the most error for each plotted subset, with lower RSS values indicating that the model  
80 accounts for more error in the data than the other model.

## 81 **Computing tools**

82 Due to R being initially formatted as a statistical language, I used R scripts for the majority of my  
83 computing, with specialist shell and python scripts for specific tasks. R's syntax made using it for  
84 data preparation, model fitting, plotting and analysis more seamless than trying to do similar tasks in  
85 Python, and R has especially superior abilities in terms of data visualisation. Despite this, Python's  
86 subprocess module made it the perfect vessel through which to run the workflow. Finally, I used bash  
87 as a script to aid in the compilation of the LaTeX code by using texcount.

## 88 **References**

- 89 [Daniel et al., 2020] Daniel, N., Casadevall, N., Sun, P., Sugden, D., and Aldin, V. (2020). The burden  
90 of foodborne disease in the uk 2018. *London: FSA*.
- 91 [Foegeding, 1997] Foegeding, P. M. (1997). Driving predictive modelling on a risk assessment path  
92 for enhanced food safety. *International Journal of Food Microbiology*, 36(2-3):87–95.
- 93 [Grijspeerdt and Vanrolleghem, 1999] Grijspeerdt, K. and Vanrolleghem, P. (1999). Estimating the  
94 parameters of the baranyi model for bacterial growth. *Food microbiology*, 16(6):593–605.
- 95 [López et al., 2004] López, S., Prieto, M., Dijkstra, J., Dhanoa, M. S., and France, J. (2004). Statisti-  
96 cal evaluation of mathematical models for microbial growth. *International journal of food microbiol-*  
97 *ogy*, 96(3):289–300.
- 98 [McMeekin and Ross, 1996] McMeekin, T. A. and Ross, T. (1996). Shelf life prediction: status and  
99 future possibilities. *International journal of food microbiology*, 33(1):65–83.
- 100 [Organization et al., 2015] Organization, W. H. et al. (2015). *WHO estimates of the global burden of*  
101 *foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015*. World  
102 Health Organization.
- 103 [Ross and McMeekin, 2003] Ross, T. and McMeekin, T. A. (2003). Modeling microbial growth within  
104 food safety risk assessments. *Risk Analysis: An International Journal*, 23(1):179–197.
- 105 [Zwietering et al., 1990] Zwietering, M., Jongenburger, I., Rombouts, F., and Van't Riet, K. (1990).  
106 Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–1881.