Report on Model Selection between Logistic Growth curves based on chosen Bacterial Population Growth Experimental Data

PokMan HO

Department of Life Sciences, Faculty of Natural Sciences,

Imperial College London

Approximate Word Count: 367

Report on Model Selection between Logistic Growth curves based on chosen Bacterial Population Growth Experimental Data

PokMan HO(CID: 01786076)

Abstract

Introduction

Multiple equations are published attempting to describe the logistic growth pattern of microbes. Limited by the research scales, these equations may not always align perfectly with all microbial growth datasets. This report is aimed at choosing one dataset from a published article and apply non-linear least square (NLLS) model selection method to compare their descriptive effectiveness on this chosen dataset.

Methods

A subset of microbial growth data was selected based on "the highest number of data points". The candidate dataset was the replicate 1 from Zwietering *et al.*(1994)[1] *Lactobaciulus plantarum* on MRS substrate under 10 degrees Celsius. The data was containing 151 records on population cell count (N).

The data was recorded in "population change" (response variable) against "time of experiment (hr)" (explanatory variable). The response variable was neither normally-distributed nor log-normal (Shapiro Test p-value: 0 ; Min 0.16 , 1st Q 3.24 , 2nd Q 6.55 , 3rd Q 7.47 , Max 8.86 , all corrected to 2 d.p.). Time (in hr) was also recorded not in a normal-distributed not log-normal way (Shapiro Test p-value: 0 ; Min 1.49 , 1st Q 57.07 , 2nd Q 123.3 , 3rd Q 195.36 , Max 345.07 , all corrected to 2 d.p.).

Expected population change calculated from Verhulst equation (classical logistic model)[2], modified Gompertz model[3], Baranyi model[4] and Buchanan model[5] were plotted against the real data on a semi-log graph of "log Population Change" vs "Time (hr)". The mechanistic models were also evaluated by NLLS method[6,7]. Pairwise comparisons of models with AIC[8,9] and BIC[10] selection methods were also carried out for identification of fittest model the the chosen dataset.

Computing tools

R (ver 3.6.0)[11] was used with following packages: "ggplot2"[12] was used for visualisation; "reshape2"[13] was used for converting dataset from wide to long format; "scales"[14] was used for improve "ggplot" graphs data presentation; and "minpack.lm"[15] was used for computing non-linear least square statistics for model comparisons.

Results

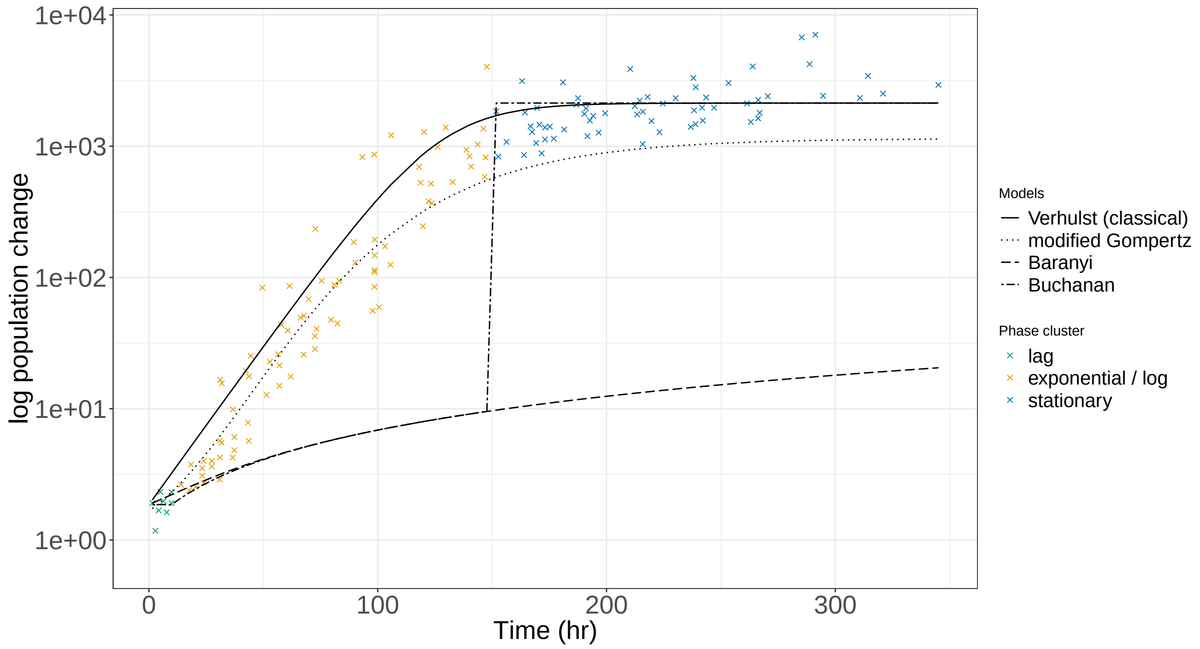

Figure 1: Semi-log graph showing four different models fitting on data of "Population Change" against "Experiment time" with points clustered into three main phases of sigmoid growth curve.

Discussion

Model fitness to real data and simplistic mathematics were favoured by both AIC[8,9,16] and BIC[16,17]. Apart from that, BIC also takes account of sample size effect[16,17].

comparisons in different fields[18–25]

Conclusion

Upon model comparisons through AIC[8,9] and BIC[10] model selection methods, XXX is the most suitable model for describing the selected dataset.

Code and Data Availability

All scripts and data used for this report were publicity available at GitHub.

References

1. Zwietering, M., De Wit, J., Cuppers, H. & Van't Riet, K. Modeling of bacterial growth with shifts in temperature. *Appl. Environ. Microbiol.* **60,** 204–213 (1994).
2. McKendrick, A. & Pai, M. K. XLV.—the rate of multiplication of micro-organisms: a mathematical study. *Proceedings of the Royal Society of Edinburgh* **31,** 649–653 (1912).
3. Gil, M. M., Brandão, T. R. & Silva, C. L. A modified Gompertz model to predict microbial inactivation under time-varying temperature conditions. *Journal of Food Engineering* **76.** Bugdeath, 89 –94. ISSN: 0260-8774. `http://www.sciencedirect.com/science/article/pii/S0260877405003389` (2006).
4. Baranyi, J, McClure, P., Sutherland, J. & Roberts, T. Modeling bacterial growth responses. *Journal of industrial microbiology* **12,** 190–194 (1993).
5. Buchanan, R., Golden, M. & Whiting, R. Differentiation of the effects of pH and lactic or acetic acid concentration on the kinetics of Listeria monocytogenes inactivation. *Journal of Food Protection* **56,** 474–478 (1993).
6. Kelley, C. T. *Iterative methods for optimization* (SIAM, 1999).
7. Strutz, T. *Data fitting and uncertainty: A practical introduction to weighted least squares and beyond* (Vieweg and Teubner, 2010).
8. Akaike, H. in *Selected papers of hirotugu akaike* 199–213 (Springer, 1998).
9. Burnham, K. & Anderson, D. Model selection and multimodel inference: a practical information-theoretic approach. *Ecological Modelling.*
10. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).

11. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2019). `https://www.R-project.org/`.

12. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* ISBN: 978-3-319-24277-4. `https://ggplot2.tidyverse.org` (Springer-Verlag New York, 2016).

13. Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software* **21,** 1–20. `http://www.jstatsoft.org/v21/i12/` (2007).

14. Wickham, H. *scales: Scale Functions for Visualization* R package version 1.0.0 (2018). `https://CRAN.R-project.org/package=scales`.

15. Elzhov, T. V., Mullen, K. M., Spiess, A.-N. & Bolker, B. *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds* R package version 1.2-1 (2016). `https://CRAN.R-project.org/package=minpack.lm`.

16. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends in ecology & evolution* **19,** 101–108 (2004).

17. Turchin, P. *Complex population dynamics: a theoretical/empirical synthesis* (Princeton university press, 2003).

18. Kuha, J. AIC and BIC: Comparisons of assumptions and performance. *Sociological methods & research* **33,** 188–229 (2004).

19. Aho, K., Derryberry, D. & Peterson, T. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* **95,** 631–636 (2014).

20. Yang, Y. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika* **92,** 937–950 (2005).

21. Penny, W. D. Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage* **59,** 319–330 (2012).

22. Vrieze, S. I. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods* **17,** 228 (2012).

23. Wang, Y. & Liu, Q. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research* **77,** 220–225 (2006).

24. Acquah, H. D.-G. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics* **2,** 001–006 (2010).

25. Edwards, D., De Abreu, G. C. & Labouriau, R. Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests. *BMC bioinformatics* **11,** 18 (2010).