

# CMEE Masters: Miniproject Assessment

February 14, 2022

**Assignment Objectives:** To address on a model-fitting problem using computational methods, and produce a written report, all in a coherent, reproducible, modular workflow under version control.

**Student's Name:** Sarah Dobson

**Overall Miniproject Mark:** 70%

## Overall Project Organization

Directories all in place and uncluttered.

You have included a `readme` file that briefly describes the project, lists the programming languages used (with version numbers), and describes the repo structure and the key files and their purposes. You do include a dependencies section, listing Rstudio and VSCode for example, however most useful would be to list packages and other things without which your project cannot run at all (whereas it can be run from terminal without Rstudio or VSCode). This would be best practise to make the life of a user of your project most painless.

Nice touch putting the Latex report source files and pdf in to their own subdirectory.

Overall a cleanly organised project. Good to see documentation, just be aware of what is most appropriate to put in it!

## The Code

Your choice of coding tools is generally appropriate, and it's fine to have a preference for a specific language like R. Do remain open to Python or C for more complex projects though. You used a sensible number of packages in your R which is good - too many packages stunt your development as a programmer and are bad for reproducibility.

Your code is reasonably commented, and simplifies the process of determining what each part of your code is doing at a glance. Your code is well partitioned into separate scripts carrying out specific functions, but we note that you have a slightly confusing approach to model fitting in which you repeatedly run `nlsLM` on each model in order to extract estimates and AICc values separately. This is somewhat inefficient, and it would be preferable to fit models a single time and save the resulting model object as a variable to then have information of all kinds extracted from it. Not only does this avoid redundant repetition of the same fits, it would also reduce the likelihood of introducing differences between model fits, as for example you appear to do by extracting AICc scores using models fit to logged population data, but then whole model estimates from fits to non-logged data. Recall also that it is considered best practise to define all necessary functions at the start of a script, rather than in an ad-hoc manner throughout the main body of the script.

We encountered one major error in trying to run your workflow, in that the `print()` statements in your master bash script should have been echo statements (`print()` is for Python, `echo` is

for bash!). Once this was fixed, the remainder of your project ran without fatal errors, which is good! You successfully fit 5 models (quadratic, cubic, logistic, Baranyi and Gompertz) to your data and compare them using AICc. However we note that, for all of your models, your AICc values were computed from fits to the logged population data. Not all of your models are usually fit to log data (the quadratic, cubic and logistic models generally are not, for instance). A better option might have been to fit these models to non-logged data, and the remaining models to logged data, and to manually calculate non-logged residuals for these so that you can still perform model comparison using AICc. Your code for Baranyi fitting also produced many (non-fatal, because you tryCatch'ed them) errors, suggesting that something about your approach to the Baranyi fitting was not working especially well.

Recall that you should write into your workflow commands that will delete all existing output files every time the workflow is run (they should be re-generated afresh). Also, put in checks so that the computational workflow aborts if any step in the analysis gives an error. Reporting that error to the user is a good idea too.

Your workflow included progress updates printed to the terminal (although they were initially miscoded as `print()` rather than `echo` statements). This is good practise as it helps the user work out how fast things are progressing and troubleshoot if needed. However it can be hard to distinguish progress updates from other terminal output while the terminal is updating frequently, so in the future consider delineating your progress updates more. For example. adding special characters makes messages stand out a lot better.

Your project ran in reasonable time ( 30s), with the most time spent model fitting. Although this is not problematically slow, adopting a more efficient coding strategy as discussed above would probably produce substantially faster runs.

Overall your project meets the project brief and achieves the model fitting and analysis that you set out to achieve. Improvements to how you choose to architect your code are likely to come with practise and from clearly planning out your code in advance and thinking carefully about how to link together each part. Your code was almost free of errors (besides the `print()`/echo confusion) and was well commented. A solid job overall.

**Marks for the project and computational workflow: 65%**

## The Report

Capable and well executed. You clearly understand the relevant concepts and how to grapple with the data using the available tools. Particularly well constructed abstract.

Title: Concise and descriptive though not especially specific.

Abstract: Excellent. Succinct background and context, clear study objectives, summary of methods and results and a take-away. (95%)

Intro: Good expansion of the background and motivation, but somewhat imprecise in places (e.g. “nonlinear models are fit to growth curves based on a minimum number of biologically relevant parameters” seems to confuse nonlinear and mechanistic models). Nonetheless, the study aims are made clear and follow quite naturally from the preceding intro text. (68%)

Methods: All essentials appear to be present. Should mention how many data subsets are left

after filtering. Models are defined clearly along with initial parameter estimates for the nlls algorithm where appropriate. Would have been nice to see AIC and AICc equations/definitions as well. Extra credit for fitting 5 models, and going beyond just AIC for comparison. (70%)

Results: Main findings clearly presented in text and with reference to Tables and Figures. Baranyi appears to have caused trouble in producing fits, only generating results for about 10 data. Consider if it's worth keeping it in given the discrepancy with all the other models. Fig 2 gives nice overview of fits across quite different data subsets. (70%)

Discussion: Key findings summarised. Reasonable discussion of differences in goodness of fit between different model types, and potential interpretations thereof with reference to the literature. Some discussion of shortcomings (mostly convergence issues) with suggested improvements. Concludes with a reasonably well distilled take-home message. A 73

(Some specific feedback is in the attached pdf, and we can also discuss more aspects of your write-up in our 1:1 feedback meeting)

**Marks for the Report:** 75%

**Signed:** Samraat Pawar & Alexander Kier Christensen

February 14, 2022

## Notes on Assessment :

- This written feedback will be discussed in a 1:1 session scheduled after this assessment has been given to you.
- The coursework marking criteria (included in this feedback at bottom) were used for both the computing and report components of the Miniproject Assessment. *In contrast*, Your final dissertation project marks are going to be based pretty much exclusively on the written report and viva (not code). Expect your final dissertation report to be marked more stringently, using the dissertation marking criteria (also included in this report).
- In the written feedback, the markers may have contrasted what you have done with what you should do in your actual dissertation. *This does not mean that you were penalized* — one of the main goals of the miniproject is to provide feedback useful for your main dissertation. However, there may be cases where what you have done is just really bad practise (for example missing line numbers or abstract), irrespective of whether it is a mini- or main- project report – you will be penalized in that case.
- The markers for this assessment are playing the role of somebody trying to understand and use your project organization and workflow from scratch. So it will seem like the feedback is particularly pedantic in places — please take it in the right spirit!