

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: Eamonn Murphy

Overall Miniproject Mark: 76%

Overall Project Organization

Your directories are all in place and free of extraneous clutter.

You have included a comprehensive `readme` file describing the project, its structure, the languages used with any necessary packages for each, and the key files and their descriptions. This is excellent programming practise and makes it much easier to run your workflow as all dependencies are clear. One minor point is that you have used a mixture of naming conventions – CamelCase (e.g. `ModelFitting.R`) and snake_case (e.g. `add_ids.py`). Although it's a purely aesthetic point, it is considered best practise to choose one convention and stick to it. Bear this in mind for future projects.

Overall the project organization and documentation are clean and logical. Good job!

The Code

The choice of coding tools is generally appropriate. It is nice to see you using a mix of R and Python for different tasks. You were sensible about the number of packages used in both R and Python, which is good – excessive reliance on packages can hinder your development as a programmer and pose problems for reproducibility.

Your R and Python code is sensibly commented, giving a clear at-a-glance sense of what is going on in each section of the code. You did leave some early “exploratory” code in `Rexplore.R`, albeit commented out. Such code can be removed in the final version of a project once you realise it is not going to be used, to avoid ending up with large blocks of code scattered alongside comments that are there to guide the reader through the useful code. Your code is generally well partitioned into different scripts, though in future, more complex projects, you might consider separating function definitions and code execution into separate scripts.

Your workflow ran error-free, which is to be commended! You successfully fit 4 models (quadratic, cubic, logistic and Gompertz) to your data, and compare them using AICc. However, although you logged the population sizes for the Gompertz and not for the rest (which is good!) you did not ensure that residuals were transformed either all into log-space or into non-log-space before computing and comparing AICc values, and your comparisons are therefore potentially biased in favour of the log-space Gompertz model.

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).

Your workflow did not post progress updates to the terminal during code execution. This would be good practise so that users can keep track progress through the workflow during execution, and can also help troubleshoot if necessary. Worth considering for future projects.

Your project ran in reasonable time, with the model fitting and initial parameter sampling taking the largest proportion of the runtime.

Overall a solid project that comfortably achieves the brief and successfully implements extras like more models and initial parameter optimisation. Cleanly organised, well commented and error-free. Nice job.

Marks for the project and computational workflow: 75

The Report

Clear understanding of the methods and tools, a commendable clarity of writing and attention to detail, evidence of analytical thinking and originality. With some restructuring and a little more meat in the results section this could have been exceptional.

Title: Concise, descriptive, presents main finding.

Abstract: Pretty good. Background is a little generic but the objectives, methods and results are clearly stated and it ends on a clear take-home message. (76%)

Intro: Expands nicely on the background/motivation suggested by the abstract, though the initial paragraph is somewhat sparsely referenced. Description of the mechanistic-phenomenological distinction is welcome since this is the focus of his study but some of the detail strays into Methods territory (particularly the tables of equations and dataset columns). The aim of the study is clearly communicated, alongside an explicit hypothesis. (74%)

Methods: Good to excellent – concise but with appropriate attention to detail where necessary. Extra credit for init parameter sampling (and for actually checking whether it was adding meaningful accuracy!). The description of the AIC/BIC/AICc model comparison stuff would ideally have been here rather than at the start of the results. Likewise the above-mentioned stuff from the Intro. Computing tools section is present. (72%)

Results: Slightly sparse considering the AIC/BIC/AICc discussion belongs in the methods. A table/summary of successful fits for each model would have been good to start off with, along with example plots of good and bad fits for some of the models (there is one such plot in the discussion). Fig 1 does nicely summarise the main finding and is appropriately referenced. (67%)

Discussion: Good. Restates the key findings and undertakes to interpret them in the context of the inherently “messy” nature of biological data, with frequent reference to the literature on the subject. Some discussion of limitations and how future studies might mitigate them is included. Concludes with a take-home message referencing Levins’ “sufficiency criteria” for model parameters, though it would have been good to explicitly state these criteria in the discussion somewhere. (78%)

(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: 76%

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!