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: An Nguyen

Overall Miniproject Mark: 67.5%

Overall Project Organization

Each directory is in place and uncluttered. You have included a concise and clearly formatted `readme` file describing the structure of the project directory, the key files and their functions. This is all very good practise. In future it would be worth considering a dependencies section, so that users can see from the readme exactly which packages they will need to run the workflow.

One minor suggestion – you could have put the writeup LaTeX source files and pdf in a separate directory – this is what you should aim to do for your final dissertation.

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

The Code

Your choice of coding tools is appropriate, and it's encouraging to see you using both R and Python for different tasks. Your package use is relatively restrained, which is good – overuse of packages limits your development as a programmer and can pose problems for reproducibility.

Your code is sensibly and concisely commented, giving an at-a-glance sense of what is happening in each section of your scripts. We note that since qpcR is ostensibly a package designed for quantitative PCR analysis (which of course does not feature in your project), it would have been helpful to include a brief comment explaining that you use this package just for the akaike.weights() function in your analysis. Your scripts are well partitioned into functions and execution. In the future (particularly for more complex projects) you might consider placing functions/modules into separate scripts from the code execution and then importing them as needed.

Your master script ran first-time without errors. This is something to be proud of! However, inspecting your code we see that you have elected to suppress all python warnings and all warnings relating to the nlsLM fitting commands in your model_fitting.R script. This is a potentially dangerous practise, since warnings can indicate when there are genuine issues with your code. In future, suppressing warnings should (at most) be something that a user can enable optionally when running your code, and if you choose to suppress warnings by default in a specific function call, you should make it clear (e.g. in the readme) why you believe warnings in that location to be worth suppressing. You successfully fit 4 models (linear, logistic, Gompertz and Baranyi) to your data, and compare them using AIC and BIC. However, we note that you chose to log the population sizes for all the models. This is an unusual choice, and technically means you have chosen to investigate whether, for example, the log of the population follows a polynomial relationship w.r.t time, rather than the population itself. A better option might have been to fit the polynomial models and logistic 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 AIC/BIC.

Recall that you ought to write into your workflow commands to delete output files every time the workflow is re-run (so that they are re-generated afresh).

You incorporate progress updates into the workflow. These are concise, informative, and you have made them stand out from the crowd of other terminal output using special characters. This is great practise, and makes it much easier to see what is happening while running your master script.

Your project ran exceptionally quickly (7s). This is partly because you elected to set fixed initial conditions rather than randomly sampling them, but it is also indicative of good programming practise and efficient code, so well done.

Overall, a very competent project. Well-organised, cleanly commented and very efficient. Some minor improvements (e.g. dependencies) and more careful treatment of warnings would be good to keep in mind for future tasks.

Marks for the project and computational workflow: 72%

The Report

You understand the methods being used reasonably well, but do not make substantial links to the wider literature, and the analysis and discussion is somewhat limited in scope and depth. Your writing is functional but suffers from occasional mistakes/awkwardness.

Title: Minimally descriptive, too vague to give any real idea of what the paper is about.

Abstract: No background, but does contain a summary of the aims, methods and conclusion. (55%)

Intro: Rudimentary background supplied, but without much reference to the literature (except for the models themselves). Brief statement of paper's aim but no specific hypothesis. Most of the intro is a description of the various models fit to the data and of AIC. The description is reasonably accurate but really belongs in the Methods. (52%)

Methods: Pretty complete (considering also the Intro material that should have come in this section). Wrangling, model fitting and model comparison are each described accurately and without extraneous detail. Computing tools section is present. (68%)

Results: Concise and clearly described. You considered a number of successful fits together with AIC and Akaike weights to arrive at an overall appraisal of each of the models. No deeper analysis o, for example, which models fit better on different subsets of the overall dataset was attempted. (65%)

Discussion & Conclusion: Somewhat brief. Key findings are stated. Discussion/interpretation of the findings is present but somewhat limited – e.g. states that Baranyi might be performing poorly on subsets with few data points, but does not evaluate whether Baranyi AIC had any relationship with sample size. Some discussion of tradeoffs between ease of model fitting and model accuracy. No discussion of the implications of the findings in the context of the wider literature. Shortcomings, where mentioned, are not accompanied by ideas for how to improve them (e.g. better model fitting through parameter sampling). (58%)

(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: 63%

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 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!