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: Uva Fung

Overall Miniproject Mark: 75%

Overall Project Organization

All directories in place, no cluttering.

You have included a `readme` file which briefly describes the project, lists the programming languages used (with version numbers), outlines the project structure, key files and their intended purpose within the workflow, and gives instructions for running the workflow. Nice job. You do include a dependencies section, listing Rstudio 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 for example). This would be best practise to make the life of a user of your project most painless.

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 a cleanly organised project with useful documentation (though packages would be ideal!).

The Code

Your choice of coding tools is appropriate, and it is fine to have a preferred language, though we would encourage you to remain open to using Python or C in future, potentially more computationally complex projects. You are reasonable in your use of R packages which is good – too many packages stunt your development as a programmer and are bad for reproducibility.

Your code is clearly and sensibly commented, giving a nice at-a-glance sense of what is going on in any given section of your workflow. Your workflow is very nicely partitioned into distinct scripts with specific tasks, but we note that you have a slightly confusing approach to model fitting in which you repeatedly run lm or 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 once and save the resulting model object as a variable to then have information of all kinds extracted from it. This avoids redundant (and thus inefficient) repetition of fits. 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.

Your workflow ran without errors, which is something to be pround of! You successfully fit several models and compare them with AIC and BIC. We note that you appear to have logged population sizes for all your chosen models. This is an unusual choice, as for example the polynomial and logistic models are generally fit to non-logged data, whereas Baryani and Gompertz

are explicitly designed for logged input data. A better option might have been to fit the polynomial 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.

Also, running your project generated warnings regarding the number of iterations in nlsLM. Ideally these should be addressed so that no warnings are given at all, but this is not always possible given time constraints, and is OK so long as the warnings aren't a sign of something fundamentally flawed.

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. 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 took some time to run (12.5 minutes), with the model fitting and initial parameter sampling taking by far the most time. Consider profiling here to see what if any time saving measures you can introduce, as for a project as complex as this it is worth making some savings. Adopting a more efficient coding strategy that avoids repeating already-executed fits (as mentioned above) would almost certainly cut down runtime for example.

Overall, a good project, cleanly organised, well commented and error-free, but with some inelegancies that slow down execution.

Marks for the project and computational workflow: 72%

The Report

Excellent abstract and intro. Methodology and results concisely and capably presented. Sensible and well reasoned discussion. Very well done overall.

Title: Concise, specific, conveys key finding.

Abstract: Excellent. Concise motivation, study objectives and methods clear. Results and conclusions succinctly summarised. (85%)

Intro: Also excellent! Motivation and background nciely expanded on with reference to the wider literature. Chosen models introduced alongside a clear and well written description of the mechanistic-phenomenological divide. Study objectives very clearly delineated into specific questions, and a preliminary hypothesis. Would have just been nice to briefly name the techniques used to evaluate and compare model fits. (85%)

Methods: All essential components present. Minor error in confusing first order polynomial model for OLS (which is a model fitting technique applicable to all linear models). Models all clearly defined, would have been nice to have equations for the model comparison metrics as

well. Good practise exhibited in checking model assumptions. Extra credit for fitting several models and implementing a search for optimal init parameter values. Computing tools present. (76%)

Results: Key findings clearly presented, with reference to Tables and Figs, though Figure text size was generally far too small. Figures give a nice sense of how different datasets exhibited different best fitting model(s). (72%)

Discussion: Key findings summarised, followed by discussion of the performance of different models and interpretation thereof. Some references made to other studies throughout but links to the literature could be expanded upon. The research questions and hypothesis spelt out in the intro are generally addressed and discussed, which is good. There is some discussion of limitations and some possible improvements mentioned. Ends on a well-distilled take-tome message. (72%)

(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: 78%

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!